DOCUMENT RESUME

ED 284 207 CS 210 434

| | |
|---|---|
| AUTHOR | Rijlaarsdam, G. |
| TITLE | Effecten van Leerlingenrespons op Aspecten van Stelvaardigheid, SCO Rapport 88 (Effects of Student Peer Feedback on Some Aspects of (Written) Composition Skills. Foundation Center for Educational Research Report 88). |
| INSTITUTION | Amsterdam Univ. (Netherlands). |
| REPORT NO | ISBN-90-6813-105-2 |
| PUB DATE | 86 |
| NOTE | 183p.; Doctor in Social Sciences Dissertation, University of Amsterdam. English summary. |
| PUB TYPE | Reports - Research/Technical (143) -- Dissertations/Theses - Doctoral Dissertations (041) |
| LANGUAGE | Dutch |
| EDRS PRICE | MF01/PC08 Plus Postage. |
| DESCRIPTORS | *Feedback; Foreign Countries; *Peer Evaluation; Peer Relationship; Secondary Education; Teacher Guidance; *Teacher Student Relationship; Teaching Methods; *Writing Evaluation; *Writing Improvement; *Writing Instruction; Writing Processes; Writing Research |
| IDENTIFIERS | Netherlands |

ABSTRACT

A study examined whether a program of writing instruction based on peer feedback would improve writing better than instruction based on teacher feedback. It was hypothesized (1) that experimental peer feedback classes would perform better than control classes on certain writing achievement variables and would display more positive attitudes toward writing and evaluation, and (2) that writing processes in the experimental classes would differ from those in the control classes. A writing program based on peer evaluation of essays was developed and refined. The control groups and the experimental groups followed exactly the same process, except that teachers in the control groups provided the feedback. A pretest, an intermediate test, and a posttest were administered; 792 essays, from pupils at eight different Dutch secondary schools on two different class levels, were evaluated. Results indicated that there was no difference between the programs on writing performance and psychological variables. Findings suggest that perhaps feedback received from peers is of lower quality than that received from teachers, or perhaps feedback is not all that important. Findings also suggest that peer feedback and teacher feedback might be in equal competition, and that intensive peer evaluation is a good, but not a better, alternative to intensive teacher feedback. (Tables of data are included, and extensive references are appended.) (NKA)

***********************************************************************
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
***********************************************************************

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED284207

# EFFECTEN VAN LEERLINGENRESPONS OP ASPECTEN VAN STELVAARDIGHEID

G.C.W. Rijlaarsdam

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Rijlaarsdam

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS210434

BEST COPY AVAILABLE

Deze uitgave werd mogelijk gemaakt door de steun van de Stichting Kohnstamm Fonds voor Onderwijsresearch.

CIP-gegevens

Rijlaarsdam, G.C.W.

Effecten van leerlingenrespons op aspecten van stelvaardigheid / G.C.W. Rijlaarsdam; (vert. summary H.S. Lake). - Amsterdam: Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam. (SCO-rapport; nr. 88)
Tevens proefschrift Universiteit van Amsterdam. - Met lit.opg. - Met samenvatting in het Engels.
ISBN 90-6813-105-2
SISO 499.5 UDC 371.012:82.08
Trefw.: stellen; onderwijsresearch.

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij electronisch, mechanisch, door fotocopieën, opnamen, of op enige manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Prijs : f 11,50 excl. porto

Te bestellen bij : Stichting Kohnstamm Fonds voor Onderwijsresearch,
Mevrouw W. Sargentini,
Grote Bickersstraat 72
1013 KS Amsterdam
tel.: 020-264547

ISBN 90-6813-105-2

Copyright: Stichting Kohnstamm Fonds voor Onderwijsresearch.

Deze onderzoekingen werden deels gesteund door de Stichting Taalwetenschap, die wordt gesubsidieerd door de Nederlandse Organisatie voor zuiver-wetenschappelijk onderzoek (Z.W.O.).

G.C.W. Rijlaarsdam

Typewerk en verzorging G.Dupré

Vertaling Summary H.S.Lake, Bussum/Amsterdam

EFFECTEN VAN LEERLINGENRESPONS OP ASPECTEN VAN STELVAARDIG-
HEID

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor in de Sociale Weten-
schappen aan de Universiteit van Amsterdam, op gezag van de
Rector Magnificus dr D.W.Bresters, hoogleraar in de Faculteit
der Wiskunde en Natuurwetenschappen, in het openbaar te
verdedigen in de Aula van de Universiteit (Oude Lutherse
Kerk, ingang Singel 411, hoek Spui),

op vrijdag 14 november 1986 om 14.30 uur

door

Gerrit Cornelis Willem Rijlaarsdam

geboren te Alblasserdam

Promotoren:

Prof. dr H.Wesdorp
Prof. dr G.J.Mellenbergh

INHOUD pagina

VOORWOORD

1. INLEIDING 1

1.1. Stelonderwijs: kritiek en problemen 1
1.2. Communicatie-onderwijs: een nieuw probleem en een oplossing 4
1.3. Opbouw van de studie 7

2. THEORETISCH KADER 9

2.1. Inleiding 9
2.2. Aanzet tot een theorie van leerlingenrespons 12
2.3. Schrijfprocestheorie 17
2.4. Onderwijskundige theorieën 20
2.5. Conclusie en hypothesen 23

3. ONDERZOEKSLITERATUUR 25

3.1. Storende invloeden bij collega-beoordelingsprocessen 25
3.2. Respons geven en krijgen op stelopdrachten 30
3.3. Effectstudies 34
3.4. Conclusies 46

4. SAMENSTELLING VAN DE EXPERIMENTELE ONDERWIJSPROCEDURE 51

4.1. Praktijkinventarisatie 51
4.2. De experimentele onderwijsprocedure: eerste versie 55
4.3. Vijf bruikbaarheidsonderzoeken 58
4.4. Definitieve experimentele onderwijsprocedure 67

5. INSTRUMENTATIE EN MEETKWALITEIT VAN DE DATAVERZAMELING 73

5.1. Opstelschalen 73
5.2. Analytische opstelscoringsvoorschriften 78
5.3. Schrijfprocesvariabelen 83
5.4. Psychologische Variabelen 87
5.5. Conclusies 90

6. METHODE EN PROCEDURES 91

6.1. Dataverzameling en toetsprocedures 92
6.2. Proefpersonen, docenten, scholen en formeren van condities 94
6.3. De onderwijsprocedures in de praktijk 97
6.4. Controles op de uitvoering van de onderwijsprocedures 105
6.5. Conclusies 115

| | pagina |
|---|---|
| 7. RESULTATEN | 117 |
| 7.1. Multivariate covariantie-analyse | 117 |
| 7.2. Resultaten Schrijfprestatievariabelen | 119 |
| 7.3. Resultaten Psychologische Variabelen | 121 |
| 7.4. Resultaten Schrijfprocesvariabelen | 123 |
| 7.5. Psychologische Variabelen en Schrijfprestaties: nadere exploraties | 126 |
| 7.6. Conclusies | 132 |
| 8. DISCUSSIE | 135 |
| 8.1. De validiteit van de conclusies | 135 |
| 8.2. De resultaten verklaard: en poging | 138 |
| 8.3. Belangwekkende nevenresultaten | 145 |
| 8.4. Consequenties voor de lespraktijk | 148 |
| SUMMARY | 153 |
| LITERATUUR | 159 |
| BIJLAGEN | |
| 1. Geraadpleegde bronnen, gebruikt bij het ontwerp van een theorie over leerlingenrespons (par. 2.2.) | 179 |
| 2. Geraadpleegde effectstudies (par. 3.3.) | 181 |
| 3. Lespraktijken op tien scholen in Nederland met leerlingenrespons (par. 4.1.) | 183 |

# 1. INLEIDING

In deze studie worden twee steldidactieken vergeleken, die verschillen op één hoofdvariabele: de feedbackgever. In de ene didactiek geeft de docent feedback op het opstel, in de andere vervullen leerlingen deze taak. Effecten van dit verschil worden verwacht op verscheidene opstelkwaliteiten, op het schrijfproces en op aan het schrijven gerelateerde psychologische variabelen. Deze studie werd ingegeven door praktische en theoretische overwegingen. De behoefte aan een beproefde steldidactiek, die gelegitimeerd is door de vakdidactiek en die praktische problemen oplost, is duidelijk aanwezig.

In deze inleiding wordt ingegaan op de kritiek die men buiten en binnen het onderwijs uit op het stelonderwijs en op de problemen waarvoor docenten Nederlands zich gesteld zien. Die problemen lijken vergroot te zijn door de verschuiving binnen het vak Nederlands naar communicatie-achtig onderwijs. Een gepropageerde en gepraktizeerde oplossing voor sommige problemen is het inschakelen van leerlingen als hulp-instructeurs in het onderwijsleerproces.

## 1.1. Stelonderwijs: kritiek en problemen

Het stelvaardigheidsonderwijs is eind 1985 weer hevig in discussie geweest na het verschijnen van enkele onderzoeksrapporten. Wesdorp onderzocht via een landelijke peiling de kwaliteit van onder meer de stelvaardigheid van twaalfjarigen, De Glopper deed dat voor vijftienjarigen (Wesdorp en Hoeksma, 1985; Wesdorp, 1985; De Glopper, 1985 a,b en c). Hoewel hun bevindingen uitsluitend van beschrijvende aard zijn, hebben de media de resultaten gretig als slecht nieuws opgepakt. De buitenwacht zag haar verwachtingen blijkbaar bevestigd. Had men niet eerder geklaagd over het peil van het schriftelijk taalgebruik? (Er wort steets meer fout gesgrefen, 1981). Had de Academische Raad niet in 1981 al geconstateerd dat het peil van de schriftelijke taalvaardigheid van studenten onder aanvaarbare normen lag en tot onderzoek opgeroepen om dat peil eens te meten? Stelprodukten zijn nu eenmaal in het oog springende resultaten van onderwijs. Of men nu wel goed leert spreken, luisteren en lezen, die resultaten van onderwijs zijn veel minder zichtbaar dan geschreven teksten. Daardoor lijken stelprodukten voor de buitenwacht te dienen als een peilstok voor een globale, algehele onderwijsevaluatie, nog daargelaten of deze produkten resultaat zijn van een gebrek aan onderwijs of van gebrekkig onderwijs.

Kritiek op het stelonderwijs is niet het alleenvertoningsrecht van de buitenwacht. Binnen het onderwijs en de vakdidactiek Nederlands is veel onzekerheid over de te volgen methode, over de beoordeling en over de functionaliteit of zingeving. De traditie van stelonderwijs = stijlonderwijs (zie bijv. Erftemeyer, 1958) is in de jaren zeventig verlaten voor

een weinig concreet uitgewerkt, maar veel aanhang verwervend vernieuwingsidee van communicatief stelonderwijs. Van stromingen of stelvaardigheidsscholen is geen sprake meer. Het lijkt erop dat er evenveel steldidactieken zijn als docenten Nederlands. Damhuis, De Glopper en Wesdorp (1983) rapporteren enquêteresultaten over steldidactische aanpakken, verkregen via 515 secties Nederlands. Een poging om via clusteranalyse tot typeringen van gepraktizeerde didactieken te komen, liep op niets uit: "Samenvattend kunnen we ten aanzien van de gestelde onderzoeksvraag concluderen dat zich in de praktijk van het Nederlandse stelvaardigheidsonderwijs geen duidelijk onderscheidbare dimensies of stromingen lijken voor te doen" (Damhuis et al., 1983, p.224). Ondersteuning voor deze conclusie levert een geheel andersoortig onderzoek. In opdracht van de toenmalige Adviescommissie voor Leerplanontwikkeling Moedertaal, werd een onderzoek verricht waarin een aantal door docenten, werkzaam in het voortgezet onderwijs, bijgehouden logboeken over hun stelonderwijs werd geanalyseerd. Ook in dat onderzoek blijkt dat men zeer uiteenlopende opvattingen heeft over wat stelonderwijs is en dat er veel variatie is in aanpakken (Verstegen, 1985).

De logboeken waren onthullend authentiek: in elk logboek kon een gevorderde N.L.O.-student een didactische misser aanwijzen. Iets dergelijks blijkt uit een zeer nauwgezet uitgevoerde etnografische studie waarin het stelonderwijs van twee leerkrachten uit het basisonderwijs sterk uitvergroot onder de loep wordt genomen (Hoogeveen en Verkampen, 1985). Hoewel er tegen dit onderzoek bezwaren te berde kunnen worden gebracht (zie Bonset, 1986 en Van Tuyl, 1986), blijkt zonneklaar dat zelfs vernieuwingsgezinde docenten gebrekkig stelonderwijs geven. Afhankelijk van wat men onder onderwijs verstaat, kan men zelfs betwijfelen of er wel van stelonderwijs sprake is.

Docenten in het voortgezet onderwijs zijn zelf ook niet erg zeker over hun stelonderwijs. In het eerder aangehaalde onderzoek van Damhuis et al. (1983) werd docenten gevraagd te melden wat hun problemen in het stelonderwijs waren. In het H.A.V.O. en het V.W.O., de schoolsoorten waarop wij deze studie richten, noemden docenten het veelvuldigst problemen die pasten in de categorieën tijdgebrek, beoordelingsproblemen en gebrekkige kwaliteit van stelprodukten, vooral voor wat betreft de inhoud en de stijl. Het zijn problemen die ook in conferenties van vakdocenten besproken worden (Verslag studiedag A.C.L.O.-M. H.A.V.O./V.W.O., 1985; Verslag studiedag A.C.L.O.-M. L.B.O./M.A.V.O., 1986). We gaan er kort op in.

Dat er beoordelingsproblemen zijn, is sinds jaar en dag bekend. Het hoogtepunt in de discussie werd bereikt toen de Commissie Modernisering Moedertaal adviseerde het opstel als examenvorm af te schaffen (C.M.M., 1977). Verhevigde aandacht in de vakliteratuur voor de storende invloeden waaraan beoordelaars bloot staan (De Groot, 1966; Wesdorp, 1981) maakten docenten gevoeliger voor de problemen. Onderzoeksinspanningen om betrouwbare procedures te ontwikkelen lijken pas nu (Blok

en Hoeksma, 1984) beloond te worden. Telkenmale blijkt weer dat docenten weinig stabiel beoordelen en onderling matig overeenstemmen in hun oordeel. Docenten weten dit van zichzelf. Wat kan men onder die omstandigheden verwachten van de kwaliteit van de feedback die leerlingen op hun opstel krijgen?

Tijdgebrek: een docent Nederlands mag zich altijd weer verheugen op volle klaslokalen. Bij een volledige baan geeft hij aan tenminste zeven klassen les. Tweehonderd leerlingen bevolken wekelijks het lokaal. Geef dan eens regelmatig een opstel op! Een geënquêteerde docent in Erftemeyer (1958, p.80): "De feitelijke toestand is echter hopeloos. Zelf gaf ik 31 uren in klassen van 25-32 leerlingen (men heeft mij hierbij niet naar mijn mening gevraagd). Het gevolg is, dat ik mijn taak op een andere manier verlicht: door aanmerkelijk minder werk te laten maken (en te corrigeren) dan ik strikt noodzakelijk vind. In mijn eigen jeugd maakten we tweemaal per maand een opstel. Ik heb daar de beste herinneringen aan, mijn geest is mede hierdoor vroeg gerijpt. Ik ontdekte mijn eigen ideeën. Ik vind het maken van opstellen, onder welbewuste leiding, zoals boven omschreven, een zegen voor jonge mensen, maar.... ik laat mijn leerlingen ten hoogste 2x per jaar een opstel maken." Het beeld is, als we afgaan op de enquêteresultaten van Damhuis et al. (1983) nog niet zo somber: er worden gemiddeld - volgens zeggen - bijna drie lesuren per maand aan stellen besteed in de derde en vierde klassen van het H.A.V.O. en V.W.O. Maar nog steeds is het opdragen van een opstel, een werkstuk of een tekst in het kader van de werkvorm Gericht Schrijven geladen met de implicatie van lezen en beoordelen. Het ter oefening opgeven van een opstel, zoals bij wiskunde sommen gemaakt worden, is geen gewoonte. Bij wiskunde kan volstaan worden met een klassikale bespreking van enige probleemtypen. Bij Nederlands wordt elk opstel gezien als een unieke prestatie die het verdient individueel gewogen te worden. Een opstel opgeven betekent een aanslag op de vrije tijd van de docent: een situatie die getuige artikelen met veelzeggende titels als Korrektie keuvel (Levende Talen, 1929), Nog eens: correctie-euvel (Houbes, 1930) en Correctie-euvel (Schneiders, 1930) niet van vandaag of gisteren is. Bovendien: het vak Nederlands is met de introductie van de communicatiedoelstelling zo breed geworden, dat ook daardoor al weinig tijd overblijft voor stelonderwijs.

Blijkbaar is de bestede tijd niet voldoende: de stelprodukten voldoen niet aan de inhoudelijke en stilistische eisen die docenten aan de betreffende jaargroep stellen. Stellen zij hun eisen te hoog? Wellicht. In elk geval ervaren docenten dat hun onderwijs niet oplevert wat zij graag zouden zien. Het is geen stimulerend werkklimaat als men geen vorderingen ziet. Als daarbij nog in ogenschouw genomen wordt dat bijna één op de vijf docenten klaagt over een gebrekkige didactiek, is de conclusie dat er veel onzekerheid is over het stelonderwijs gerechtvaardigd.

Dan te bedenken dat men weet dat de correctie en beoordeling van opstellen in veel gevallen lood om oud ijzer is. Een verplicht nummer dat men snel afwerkt: snelle feedback lijkt in het algemeen verkieslijk te zijn en docenten zeggen dan ook de opstellen 4 à 7 dagen na het volbrengen van de opdracht te retourneren (Damhuis et al., 1983). Maar leerlingen kijken naar het cijfer en zijn niet te beroerd het opstel nog voor het einde van de les richting prullenbak te werpen. Onderzoeken waarin verschillende wijzen van docentfeedback met elkaar geconfronteerd werden, toonden aan dat bijvoorbeeld een cijfer of zeer uitvoerig commentaar geen significante verschillen bewerkstelligden (Fellows, 1932, geciteerd bij Pierson, 1967); marginaal commentaar versus uitgebreid commentaar lijkt evenmin verschil te maken (Bata, 1972); of de docent elk opstel van een leerling nakijkt of af en toe maar eens eentje uit een verzameling, lijkt niet uit te maken (Wagner, 1975). De docent die opstellen opgeeft en de tas met te corrigeren opstellen op de bagagedrager bindt, weet dit wel, maar voelt zich toch moreel verplicht alle opstellen zo spoedig en nauwkeurig mogelijk te corrigeren. Gedenk hem.

## 1.2. Communicatie-onderwijs: een nieuw probleem en een oplossing

Toch krijgt de steldidactiek aandacht van vakdidactici en geïnstitutionaliseerde leerplanontwikkeling (Stichting voor de Leerplanontwikkeling). De gedachte het examenopstel te vervangen, leidde tot een pleidooi voor het gedocumenteerde opstel. Een commissie werkte dit idee uit (Eigen Zorg, 1971; 1972; 1973) en droeg haar werkzaamheden over aan de Commissie Modernisering Moedertaalonderwijs. Deze initieerde het leerplanontwikkelingsproject Gericht Schrijven. Daarin werden de door Drop (1971) geïntroduceerde begrippen doel- en publiekgerichtheid, naast verifieerbaarheid, als uitgangspunten opgenomen. Communicatieprocessen èn informatieverwerkingsprocessen werden de pijlers van het nieuwe curriculum Nederlands. Dat veranderde niet toen de S.L.O. de bemoeienis met leerplanontwikkeling van de C.M.M. overnam.

De vakdidactiek Nederlands stimuleerde de opvatting dat het vak Nederlands een communicatievak was. De vakdidactiek ontwikkelde zich in de zeventiger jaren tot een zelfstandige discipline binnen (sommige vakdidactici zouden liever zeggen: buiten) de Neerlandistiek. Handboeken als Zeggenschap (Griffioen en Damsma, 1978), Tegenspraak (Griffioen, 1982), The Complete Mothertongue Curriculum (Ten Brinke, 1976), Moedertaaldidactiek (Leidse Werkgroep, 1980), Taaldidactiek aan de basis (Nijmeegse Werkgroep, 1978), alle zien de ontwikkeling van communicatievaardigheid als de wezenlijke doelstelling van het vak. Kroon (1985) geeft een historisch overzicht van het decennium 1970-1980 voor wat het denken over het vak Nederlands betreft. Kroon presenteert tevens enquêteresultaten die laten zien dat veel docenten communicatie-achtig onderwijs wel

sympathiek vonden en voorzagen dat de communicatie-tendens zou doorzetten, maar dit soort onderwijs nog nauwelijks of niet praktizeerden.

Voor het stelonderwijs had dit communicatie-onderwijs dankzij de inspanningen in het S.L.O.-project Gericht Schrijven (De Zanger, 1981) wel vormgekregen. In leerboeken als Opbouw (Bult et al., 1979) en Functioneel Nederlands (Biegstraaten et al., 1977), - zij bestrijken een groot deel van de markt -, wordt in de derde klas een flink aantal lessen besteed aan informatie verzamelen, bronnengebruik, doelbepalen, publiek bepalen etc. Met die introductie van het communicatieve aspect in leerboeken, kwam sporadisch soms en vanuit verschillende intenties, de rol van de leerling in een ander daglicht te staan. Groepswerk werd geïntroduceerd, onder meer om de docent minder teksten te laten nakijken (De Zanger, 1981). Leerlingen worden expliciet aangemoedigd in sommige leerboeken om elkaars teksten te lezen. In sommige methoden wordt dit opgedragen om inductief leren te stimuleren: eerst schrijven naar aanleiding van een opdracht, dan al lezend ontdekken aan welke eisen een goede tekst moet voldoen (zie bijv. Tussen de Regels, deel 2 van Meijerink en Vos, 1978, en Functioneel Nederlands, deel 3VH van Biegstraaten et al., 1977; Taal Vandaag van Vos en Van der Laan, docentenhandleiding, 1978). In andere methoden is het uitgangspunt niet zozeer leertheoretisch, alswel pragmalinguïstisch. In Opbouw, deel 3 HV (Bult et al., 1979) wordt nagestreefd de in de schrijfopdracht vermelde lezers hun reactie op de geschreven tekst te laten formuleren. Meestal vermeldt men als publiek in de schrijfopdracht de medeleerlingen, klasgenoten, maar het kunnen ook brugklassers zijn voor wie geschreven wordt en wier reactie gevraagd wordt. Doel hiervan lijkt communicatieve effectiviteitsbeleving. In Over en Weer (Feteris en Lewis, 1983) wordt bijna elke schrijfopdracht vergezeld van de toevoeging de tekst van een andere leerling met een analytisch schema te beoordelen: effectiviteitsbeleving en het ontwikkelen van een gevoel voor een goede tekst lijken hier samen te gaan.

Vlasman, die zelfstandig de enige complete steldidactiek voor het Nederlandse taalgebied ontwikkelde (De papieren wereld, z.j.) noemt als argumenten om leerlingen in te schakelen de taakverlichting voor de docent en de verhoging van de schrijfmotivatie. Daarnaast hanteert hij een leertheoretisch argument. Hij verwacht een effect van 'modeling': van goede en slechte teksten kunnen leerlingen leren over hun eigen aanpak en hun eigen produkten.

Vakdidactische handleidingen noemen veelal de mogelijkheid leerlingen als beoordelaar in te schakelen. Het begon reeds in de nu ten onrechte verguisde Handleiding van Van Dis (Van Dis, z.j., p.57): 'Voorts kan men de leerlingen in alle klassen nu en dan gelegenheid geven om fouten en weinig sierlijke gedeelten in elkaars werk te zoeken, voordat de leraar het werk onder ogen neemt. Als aldus gevonden fouten niet 'meetellen'

en de leerlingen elkaars punten omhoog kunnen brengen, zal een verbeten foutenjacht ontstaan, wat ook nog ten goede zal komen aan de belangstelling van de leerlingen voor de bespreking."
In meer recente handboeken speelt het communicatieve aspect een doorslaggevende rol. Ten Brinke (1976) pleit voor reageren in plaats van beoordelen. In Moedertaaldidactiek (1980), werkt Bonset Ten Brinkes ideeën uit: de leerling krijgt een reactie van een echt publiek, net als in het leven buiten de school. In Taaldidactiek aan de basis (1978) neemt de argumentatie een andere wending: voor een kind is het moeilijk afstand te nemen van de eigen tekst en zich te verplaatsen in de rol van de lezer; daarom is het aan te bevelen dat het de hulp inroept van een medeleerling. Het meest uitgewerkt en consequent doorgevoerd is het inschakelen van leerlingen in Tegenspraak (Griffioen, 1982). Ook in hun handboek Zeggenschap (Griffioen en Damsma, 1978) was de werkwijze al opgenomen, maar in Tegenspraak zijn de achterliggende ideeën consistenter. Griffioen bepleit het instellen van schrijfgroepen, bestaande uit drie à vier leerlingen, die twee doelen dienen: gehoor te zijn en feedback te geven. Zijn idee is dat schrijven zo werkelijkheidsgetrouw mogelijk moet zijn: er moet een publiek zijn. Hij combineert de pragmalinguïstische consequentie (gehoor te zijn) met een leertheoretische: vaardigheden verwerft men door activiteiten te verrichten en op de uitvoering ervan terug te zien. In zijn termen: doen + nadenken over dat doen (Griffioen en Damsma, 1978). Feedback geven biedt de gelegenheid te reflecteren over communicatie en daardoor te leren.

Als een in het begin van de zeventiger jaren opgeleide docent Nederlands, is schrijver dezes ook besmet geraakt met het communicatievirus. Liet de vaksectie aan zijn school eerst literaire werkstukken schrijven, later ging men ertoe over leerlingen voor het schoolonderzoek een artikel over een maatschappelijk onderwerp te laten schrijven. Men vond dat zinniger. Leerlingen moesten een betogend stuk schrijven voor medeleerlingen. Die adressering stelde de docenten in de gelegenheid leerlingen te verplichten hun publiek te analyseren. Op een gegeven moment begon men problemen te onderkennen met de beoordeling van de artikelen: zij waren geschreven voor leerlingen en de docenten behoorden niet tot dat publiek. Zij konden zich niet verplaatsen in de leessituatie van het publiek, zoals Griffioen en Damsma (1978) dat voorschreven. De volgende stap was snel gezet. Als experiment beoordeelden leerlingen in drie H.A.V.O.-5 klassen in het schooljaar 1978-1979 ieder drie artikelen van medeleerlingen geschreven in het kader van het schoolonderzoek. De docent beoordeelde ook: zijn cijfer telde voor 50% mee. De gang van zaken en enkele psychometrische gegevens zijn gerapporteerd in Blok en Rijlaarsdam (1981) en Rijlaarsdam en Blok (1981). Uit aan het eind van dat cursusjaar geschreven 'learner reports' bleek dat leerlingen de activiteit leerzaam vonden. Dat bracht de vaksectie ertoe

leerlingen niet pas aan het eind van hun schoolcarrière in de gelegenheid te stellen iets te leren: het becommentariëren van elkaars werk werd een constante in de lessen Nederlands. Gezien de huidige leerboeken, de vakdidactische handleidingen en de enquêtegegevens van Damhuis et al. (1983) was die sectie niet uniek in de ontwikkelingsgang. Toen bleek dat de overwegingen voor deze didactische ingreep zinnig waren en de werkwijze zo algemeen geadviseerd werd, rees de onderzoeksvraag: schrijven leerlingen die elkaars teksten becommentariëren beter opstellen dan leerlingen die feedback van hun docent krijgen?

## 1.3. Opbouw van de studie

Het onderzoek bestond uit drie onderdelen. Er werd ten eerste een experimenteel stelonderwijsprogramma voor de derde klas van het H.A.V.O. en het V.W.O. ontwikkeld, er werden ten tweede instrumenten geconstrueerd voor het meten van de afhankelijke variabelen en er werd ten derde een experiment uitgevoerd.

De hoofdstukken 2 (Theoretisch kader) en 3 (Onderzoeksliteratuur) vormen de basis voor die drie onderzoeksactiviteiten. In hoofdstuk 4 wordt beschreven hoe de experimentele stelonderwijsprocedure ontwikkeld en beproefd is. In hoofdstuk 5 wordt de instrumentatie beschreven voor drie groepen variabelen: Schrijfprestatievariabelen, Schrijfprocesvariabelen en Psychologische Variabelen. Hoofdstuk 6 is gewijd aan de uitvoering van het experiment in 22 klassen. Hoofdstuk 7 bevat de resultaten van het onderzoek. De studie besluit met een bespreking van de resultaten van het onderzoek (hoofdstuk 8).

Bij dit rapport behoren twee bijlagenrapporten, te verkrijgen bij de Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam:

G.C.W. Rijlaarsdam, Effecten van leerlingenrespons op aspecten van stelvaardigheid. Bijlagenrapport 1: Het experimentele stelonderwijsprogramma. Leerlingenboek en docentenhandleiding. S.C.O. Amsterdam, 1986.

G.C.W. Rijlaarsdam, Effecten van leerlingenrespons op aspecten van stelvaardigheid. Bijlagenrapport 2: Instrumenten. S.C.O. Amsterdam, 1986.

# 2. THEORETISCH KADER

## 2.1. Inleiding

De theorie van de moedertaaldidactiek staat nog in de kinderschoenen. Ook het theoretisch gehalte van de didactiek van het stellen is laag. Systematische doordenking van de wijze waarop schrijfvaardigheid zich ontwikkelt, gebonden is aan de cognitieve ontwikkelingsfasen van leerlingen en van de wijze waarop didactische maatregelen invloed kunnen uitoefenen op de stelvaardigheidsontwikkeling, heeft nog niet erg intensief plaatsgevonden.

Wat geldt voor de didactiek van het stellen in het algemeen, geldt ook voor één specifieke methode: het inschakelen van leerlingen als lezers en beoordelaars van elkaars teksten, ofwel leerlingenrespons. Dé theorie van leerlingenrespons bestaat niet. Wat wel bestaat is een bonte verzameling praktijkopvattingen: ideeën, legitimeringen en ervaringen van praktisch handelende docenten. Toch zijn er voor wat leerlingenrespons betreft theoretische bijdragen: in de publikaties van Moffett (1968), Elbow (1974) en Bruffee (1980) worden respectievelijk drie theorieën gepresenteerd over schrijfonderwijs waarin leerlingenrespons een grote rol speelt. Deze theorieën zijn evenwel niet empirisch getoetst.

Omstreeks 1970 verschenen in de Verenigde Staten twee publikaties die een zich in de praktijk ontwikkelende didactiek een theoretische fundering gaven. Ik doel hier op de boeken van Moffett (1968) en Elbow (1974). Een derde publikatie, die van Bruffee (1980), verscheen later, maar zijn ideeën werden vanaf 1970 in tijdschriften gepubliceerd (Bruffee, 1975). Deze drie auteurs pleiten voor het inschakelen van medeleerlingen en -studenten bij het bevorderen van elkaars stelvaardigheidsontwikkeling, zij het alle drie met andere theoretische accenten.

Moffett (1968) onderbouwt zijn pleidooi voor leerlingenrespons vanuit zijn moedertaalleertheorie, die er kort gezegd op neerkomt dat men spreken en schrijven al doende, met vallen en opstaan leert. Feedback op de spreek- of schrijfhandeling is in zulk trial-and-error leren van eminent belang, want alleen dan kan de handelende persoon te weten komen dat en waarom hij viel, en een strategie bedenken en uitproberen waarmee hij zich staande zou kunnen houden. Moffett (1968) pleit ervoor de feedback te laten geven door degeen die echt in de communicatieve handeling betrokken is: degeen voor wie het spreken en het schrijven bedoeld is. Het verdient de voorkeur leerlingen voor elkaar te laten schrijven, omdat ze dat graag doen, en dan zijn het dus ook die leerlingen die feedback moeten geven.

Elbow (1974) onderschrijft de noodzaak van feedback bij het leren schrijven. Zijn theorie is eerder communicatiepsychologisch dan taalleertheoretisch. De centrale gedachte van Elbow is dat de gangbare steldidactiek gebaseerd is op een

theorie over de kwaliteit van een goede tekst. Elbow ontkent echter het bestaansrecht van zo'n theorie. Generalisaties over functies van tekstkenmerken zijn niet mogelijk, omdat succes van het functioneren van die kenmerken afhankelijk is van de individuele lezer. Hij acht het van belang dat schrijvers leren hoe lezers teksten lezen en ervaren. Kennis van lezersreacties functioneert in een volgende schrijftaak als een - natuurlijk feilbaar - peillood. Elbows didactiek is dan ook gericht op subjectieve lezersreacties: elke reactie is legitiem, omdat niemand kan zeggen of de reactie correct is of niet. Elke reactie is waardevol, omdat de schrijver inzicht krijgt in individuele leesprocessen.

Bruffee (1980; 1981) legitimeert zijn keuze voor een steldidactiek met leerlingenrespons met kennistheoretische ideeën. Objectieve kennis is volgens hem een fictie. Kennis is subjectief: leerlingen maken steeds keuzen in het informatie-aanbod en verwerken de binnengekomen informatie individueel tot verschillende bestanden. Schrijven is een vergelijkbaar proces: een schrijver neemt al schrijvend honderd-en-één beslissingen. De basis voor die beslissingen wordt gevormd door de aanwezige kennis. Maar, betoogt Bruffee, die kennis is subjectief. Zekerheid over de correctheid van beslissingen heeft een schrijver dientengevolge nooit. Om nu inzicht te krijgen in de subjectiviteit en om de beperktheid van het subjectieve kenniskader te overstijgen, is samenwerking tussen schrijvers een probaat middel. Dit samenwerken kan verschillen van samen over een onderwerp praten tot kritieken op elkaars teksten schrijven: centraal staat het uitwisselen van inhouden om de kijk op de wereld die in teksten wordt uitgedrukt te confronteren met andere perspectieven, te nuanceren en te leren kennen.

Deze drie pleitbezorgers van leerlingenrespons staan niet alleen. In de Verenigde Staten verschijnen in tijdschriften voor leraren praktische uitwerkingen en praktijkverslagen waaraan meestal ook theoretische overwegingen toegevoegd worden. Soms steunen de artikelen expliciet op Moffett (1968), Elbow (1974) of Bruffee (1980), maar vaker nog is er sprake van een zelfstandig ontwikkelde praktijk'theorie' (zie ook Rijlaarsdam en Blok, 1981). De kernvraag van zo'n praktijk-'theorie' is: "Waarom laat je leerlingen X doen?" En het kernantwoord: "Omdat ik denk dat het goed voor hun Y en Z is." Zo is elke docent een pleitbezorger.

Het zijn theorieën zonder wetenschappelijke empirische toetsing, geen waarheden, maar vooral denkbeelden. Al rationaliserend proberen praktisch handelende docenten hun observaties met elkaar in verband te brengen. Steun voor hun theoretische constructies zullen ze niet vinden bij empirisch getoetste didactische theorieën: er zijn geen fundamentele stelvaardigheidsleertheorieën die omgezet kunnen worden in didactische, praktische toepassingen. Uitgangspunt van didactisch handelen is de praktijk: déze leerlingen en déze school en dít leerboek. Docenten toetsen dag in dag uit hun eigen opvat-

tingen over hoe het leren van moedertaalvaardigheden verloopt, welke didactische aanpak successen boekt. Dat zij de eigen praktijksituatie niet kunnen ontstijgen en dat zij opgesloten blijven in hun eigen denkkader, zodat ze alleen observeren wat hun theorie hun ingeeft, hindert hen niet. Misschien konden Damhuis, De Glopper en Wesdorp (1983) daarom zo weinig lijn ontdekken in de Nederlandse praktijk van het stelonderwijs. Iedere docent heeft een eigen proeftuin.

Wij koesteren niet de illusie dat onderwijskundig, empirisch onderzoek leidt tot 'fool-proof' didactieken, die, mits zorgvuldig gevolgd door docenten, tot resultaten zullen leiden. Daarvoor zijn er teveel variabelen in de onderwijsleersituatie. Ook onderzoek op heel fundamenteel niveau is op dit moment voor de onderwijspraktijk minder geschikt, omdat de afstand tussen experiment en onderwijsleersituatie groot is en dientengevolge resultaten uit dergelijk onderzoek door practici niet geïmplementeerd worden. Wat momenteel wel mogelijk is, theoretisch en praktisch, is onderzoek van zogenaamd molair karakter (Cook en Campbell, 1979, p.32 e.v.) waarin didactieken die op hoofdkenmerken verschillen, met elkaar vergeleken worden. Het molaire karakter van zo'n onderzoek heeft tot gevolg dat niet exact bepaald kan worden waardoor het positieve effect van een didactiek veroorzaakt wordt. Als twee steldidactieken met elkaar vergeleken worden, één gebaseerd op leerlingenrespons en één op docentfeedback, dan is bijvoorbeeld niet te bepalen of het effect op het conto te schrijven is van de hoeveelheid feedback, de aard van de feedback of het lezen van elkaars opstellen. Maar dit betekent allerminst dat dergelijk onderzoek theoretisch en praktisch niet interessant is. Ten eerste worden vooronderstellingen en goede bedoelingen, waarop de didactiek gebaseerd is, en, inherent aan haar handelingskarakter, gebaseerd zal blijven, getoetst: zijn zij fundament of drijfzand? Empirisch onderzoek is noodzakelijk omdat de rapportages van docenten geen doorslaggevend bewijs voor het effect van een didactische ingreep zijn: er zal allicht sprake zijn van een neiging tot rechtvaardiging van de inspanningen (zie Meuffels, 1982, p. 6). Ten tweede biedt een dergelijk onderzoek de praktijk beproefde en getoetste didactieken. Theoretische verfijning, het vaststellen van de precieze oorzaak die tot het gewenste leereffect leidt, kan later nog plaatsvinden.

Aan het in deze studie gerapporteerde onderzoek naar effecten van leerlingenrespons op aspecten van stelvaardigheid, lagen zekere theoretische inzichten ten grondslag. Via abstractie en analyse van praktijkrapportages schetsen we een theorie van leerlingenrespons: welke kenmerken van die didactiek zijn oorzakelijk verbonden met welke effecten? Daartoe analyseerden we literatuur over het inschakelen van leerlingen in het stelonderwijs. De voor het merendeel via een computersearch verkregen literatuur bestond uit artikelen waarin docenten van hun praktijk getuigen en die praktijk propageren. Deze literatuur werd geanalyseerd op uitspraken over waarge-

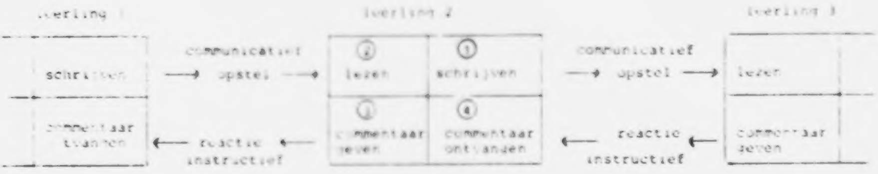
nomen of vermoede effecten. Tevens werd waar mogelijk aangetekend aan welke fase van het onderwijsleerproces men een genoemd effect toeschreef. Op basis van deze praktijkuitspraken - mengeling van werkelijkheid en retoriek -, wordt in de volgende paragraaf een aanzet tot een theorie van leerlingenrespons ontwikkeld. We gaan daarbij voorbij aan beperkingen en mogelijke nadelen van leerlingenrespons (vgl. Rijlaarsdam en Blok, 1981). Daarna wordt deze theorie ingepast in het model van het schrijfproces van Hayes en Flower (1980, 1983) (par. 2.3) en in onderwijskundige theorieën die in verband met peer evaluation geformuleerd zijn (par. 2.4). Dit hoofdstuk wordt besloten met conclusies en hypothesen.

## 2.2. Aanzet tot een theorie van leerlingenrespons

Het gehele proces van leerlingenrespons, toegepast in het stelonderwijs, laat zich grofweg ontleden in vier fasen: 1. schrijven; 2. lezen; 3. commentaar geven; 4. commentaar ontvangen. Deze vier fasen vormen twee aan twee complementaire koppels. In de eerste twee fasen, het communicatieve koppel schrijven en lezen, is de communicatieve handeling dominant: van intentioneel leren is nauwelijks sprake (zie Rijlaarsdam en Hulshof, 1984, p. 190). In het tweede - instructieve - koppel staan twee leerlingen wederom in een communicatieve relatie tot elkaar, maar nu in de rollen van onderwijsgever - onderwijsontvanger. Er wordt een instructieve boodschap verstuurd: de respons of het commentaar op het opstel.

Op elk moment vervult een leerling tenminste twee rollen in dit onderwijsleerproces. Als hij schrijft, is hij ook publiek voor de tekst van een ander. Als hij leest, leest hij als lezer, maar ook als leerling; hij leest ook om ervan te leren. Als hij commentaar geeft, is hij ook de geadresseerde voor een andere leerling. Als hij commentaar ontvangt, is hij voor een andere leerling commentator. Hieronder worden in figuur 2.1. deze relaties in beeld gebracht.

Figuur 2.1.: Een schema van leerlingenrespons



Leerling 2 vormt met leerling 3 en met leerling 1 een communicatief koppel: hij schrijft zijn opstel voor leerling 3, hij is lezer van het opstel van leerling 1. Met dezelfde leerlingen vormt hij daarnaast een instructief koppel: hij vervult de rol van docent als hij leerling 1 commentaar geeft en de rol van leerling als hij commentaar krijgt van zijn 'docent', leerling 3.

De uitspraken in de geraadpleegde literatuur werden in deze vier fasen van leerlingenrespons gecategoriseerd. Het resultaat van deze analyse van praktijkuitspraken volgt hieronder. Het vermelden van de vindplaatsen van de uitspraken hebben we achterwege gelaten. Het zou leiden tot reeksen bronvermeldingen achter welhaast elke zin en zou licht de indruk wekken dat het om een zeer gedocumenteerde theorie gaat. Zover is het niet. Wat volgt is een poging tot abstractie van praktijkuitspraken en -opvattingen. In bijlage 1 zijn de geraadpleegde bronnen vermeld.

Schrijven. De taaksituatie van de schrijver wordt bij leerlingenrespons door drie elementen gekenmerkt: de te schrijven tekst zal (1) vooral gelezen worden (en niet beoordeeld) door (2) verscheidene (3) medeleerlingen. De situatie wordt zo een communicatieve situatie. Dit in tegenstelling tot de situatie waarin de docent zal oordelen: leerlingen zullen deze meer als een evaluatieve onderwijsleersituatie ervaren. Aan elk van deze drie elementen worden effecten op het stellen toegeschreven.

Dat de nadruk op lezen ligt en minder op beoordelen, al is in de leeshandeling een oordeelvorming geimpliceerd, maakt de situatie voor de schrijvers minder bedreigend. Ze merken dat er interesse bestaat voor wât ze schreven en niet alleen voor de prestatie die ze leverden. Het zelfvertrouwen zou daardoor toenemen en leerlingen zouden vaker gaan schrijven, ook voor zichzelf. Het tweede kenmerk van de taaksituatie, de veelvuldigheid van lezers, zou ertoe leiden dat schrijvers zorgvuldiger gaan formuleren en redigeren. Het belang dat schrijvers gaan hechten aan de inhoud van hun boodschap, brengt hen ertoe aandacht te schenken aan helderheid, expliciete formuleringen, congruentie tussen wat ze bedoelen te zeggen en wat ze zeggen. Er vindt dus meer reflectie plaats onder invloed van het element publiek. De toegenomen inzet voor een goede prestatie kan ook veroorzaakt worden doordat meer lezers kennis zullen nemen van de tekst en het oordeel over de tekst de schrijver daardoor meer raakt.

Een dergelijk effect kan ook veroorzaakt worden door de persoon van de lezer. Juist ten overstaan van medeleerlingen willen leerlingen niet falen. Dat de tekst voor medeleerlingen bedoeld is, maakt het de schrijvers gemakkelijker inhoudelijk iets nieuws mee te delen in hun tekst. Het schrijven zou gemakkelijker verlopen omdat leerlingen 'gewonere' eisen aan een tekst stellen dan de vak-eisen van de docent. De inhoud zou authentieker worden.

Lezen. Leerlingen krijgen plezier in schrijven, doordat zij goede teksten lezen en zelf betere teksten gaan schrijven. Verder neemt het vertrouwen toe in eigen kunnen voor wat schrijven betreft. Leerlingen krijgen door het vergelijken van hun eigen prestatie met die van anderen meer inzicht in hun kunnen. Ze worden zich ook meer bewust van het eigene van hun manier van schrijven. De wetenschap dat andermans teksten ook niet volmaakt zijn, kan een steun zijn.

Leerlingen doen allerlei kennis op die zij in andere schrijftaken kunnen gebruiken. Ten eerste breiden leerlingen hun voorraad aanpakken, tekstmodellen en hun taalschat uit: stijlen, mogelijkheden om het publiek aan te spreken, vocabulair etc. Door teksten van hun publiek te lezen, leren ze hun publiek kennen. Nu kan men aanvoeren dat het idee van goed voorbeeld doet goed navolgen al langer in de steldidactiek opgang maakt. Didactici die de didactiek van leerlingenrespons voorstaan, wijzen echter op een belangrijk motiverend element in deze didactiek. Leerlingen lezen nu namelijk teksten die geschreven zijn door medeleerlingen, waarvan een doeltreffender modelwerking zou uitgaan dan van teksten van gevestigde schrijvers die zo vaak in (Amerikaanse) lesboeken worden opgenomen. Het gaat in leerlingenrespons om teksten die het resultaat zijn van eenzelfde opdracht als de lezende leerlingen zojuist hebben uitgevoerd, zodat de vergelijkingsmogelijkheden voor de hand liggen en leerlingen kunnen bepalen welke teksten nu op te vatten zijn als geslaagde proeven. Zeker de goede teksten motiveren tot kennisverwerving en navolging, omdat leerlingen het een prettig idee vinden dat medeleerlingen in staat zijn tot zulk goed werk. Ten tweede ervaren leerlingen de dynamiek van communicatie en kennis. Zo zien zij dat schrijftaken op een heel verschillende wijze aangepakt kunnen zijn, met succes. Veel oplossingen van het schrijfprobleem zijn goede oplossingen. Door dit inzicht zouden leerlingen meer durven te divergeren. Verder beseffen leerlingen dat zij als lezers hun eigen referentiekaders in het geding brengen als zij teksten van medeleerlingen lezen en die teksten becommentariëren. Het gevolg zal zijn dat zij in hun schrijversrol zullen beseffen dat lezers hun individuele kennis, waarden en normen betrekken bij het lezen. Dit inzicht geldt ook de relatie tussen schrijver en lezer: een tekst is een subjectief ontworpen geheel van betekenissen, iets wat leerlingen veel minder sterk ervaren als het om teksten uit leerboeken gaat. Daardoor zouden leerlingen meer verantwoordelijkheid gaan voelen voor wat ze schrijven, zeker door teksten in de openbaarheid te brengen. Het lezen van andermans teksten leidt ertoe dat leerlingen hun kennis over thema's uitbreiden en in staat worden gesteld de kijk op de werkelijkheid die zij in hun tekst schiepen te corrigeren.

Doordat leerlingen veel teksten van medeleerlingen moeten lezen, wordt hun een flinke hoeveelheid leesoefening geboden. Hierdoor worden ze perceptiever, ook met betrekking tot de eigen tekst. Fouten in andermans teksten worden gemakkelijker

onderkend dan in de eigen teksten. Maar de transfer naar toepassing op eigen teksten kan plaatsvinden omdat leerlingen in de loop van de lessen de eigen zwaktes hebben leren kennen, waardoor leerlingen meer gemotiveerd zijn om kritisch naar het eigen werk te kijken. Gevolg van de toegenomen leesvaardigheid is een intensiever en zorgvuldiger corrigeren en herschrijven.

Commentaar geven. Door het becommentariëren van elkaars teksten, ontwikkelen leerlingen die vaardigheid. Ze ontwikkelen een metataal, leren te praten en te schrijven over hoe zij een tekst waarnemen en leren hun oordeel aan anderen mee te delen. Deze commentaartaak is een heel reële communicatietaak en zou naar andere schrijftaken transfereren.

Door het geven van commentaar op elkaars teksten gebruikt de leerling impliciet of expliciet criteria voor een goede tekst. Deze kennis van criteria die leerlingen verwerven en toepassen in de responstaak, zal ook gebruikt worden als leerlingen een schrijftaak uitvoeren. Kennis van criteria leidt ertoe dat leerlingen de respons die zij op de eigen teksten krijgen beter begrijpen.

De responstaak stelt leerlingen in staat de eigen teksten intensiever te beschouwen; zij hebben geleerd eisen te stellen aan de teksten van medeleerlingen en gaan daardoor ook eisen stellen aan de eigen teksten die zij dan ook kritischer en afstandelijker lezen. Een belangrijke voorwaarde daarvoor is het afstand kunnen nemen van een tekst: de schrijver moet uit de relatie schrijver-tekst kunnen stappen. Een van de grote winstpunten van leerlingenrespons is nu juist, dat leerlingen zich los kunnen maken van een tekst en deze vanuit een meer objectief standpunt kunnen lezen. Leerlingen gaan dan ook meer rekening houden met te verwachten reacties van lezers. Er is meer aandacht voor de congruentie tussen wat de schrijver bedoelt en wat er in de tekst staat.

Commentaar ontvangen. De feedbacksituatie bij leerlingenrespons verschilt op drie kenmerken van de traditionele situatie waarin leerlingen feedback van een docent krijgen: het aantal feedbackgevers, de snelheid van de feedback en de persoon van de feedbackgever. Het grotere aantal feedbackgevers leidt ertoe, dat leerlingen minder afhankelijk zijn van één oordeel. Zij moeten zelf uit de diversiteit van feedbackboodschappen een keuze maken.

Elbow (1974) vindt het van belang dat de schrijvers van veel lezers reacties krijgen, opdat schrijvers individuele leesprocessen kunnen ervaren of reconstrueren. Moffett (1968) acht het inschakelen van meer feedbackgevers van belang omdat het de kans vergroot dat niet alleen idiosyncratische trivialia gerespondeerd worden. De ontvangers ontdoen de feedback van de merkwaardigheden van respondenten. Het blijkt schrijvers dat er een consensus te reconstrueren is uit de feedback van de lezers over sommige aspecten die zij als objectief gaan

beschouwen. Schrijvers identificeren daarnaast aspecten die individuele waardeoordelen vereisen. Vergroting van het aantal commentaren heeft voor schrijvers het voordeel dat de feedback onpersoonlijker is, minder toe te schrijven is aan eigenaardig lees- en reageergedrag van één lezer. Het wordt de schrijver daardoor gemakkelijker gemaakt de feedback ter harte te nemen.

De snelheid van feedback is een tweede aspect van de feedbacksituatie bij leerlingenrespons dat tot een positief effect op de schrijfvaardigheid leidt. De kans dat schrijvers tot het inzicht komen dat schrijven communiceren is, lijkt groter te worden als de tijdspanne tussen schrijven en respons korter is. De overeenkomst met de spreeksituatie wordt sterker naarmate de feedback sneller plaatsvindt. Schrijvers worden ogenblikkelijk de ogen geopend voor wat zij nu op papier hebben gezet. Zeker als de lezers luisteraars worden, in de situatie waarin schrijvers hun tekst voorlezen, wordt de feedbacksituatie, - luisteraars reageren fysiek! -, heel gelijkend op die van een spreek-luistersituatie.

De persoon van de feedbackgever zorgt ervoor dat de feedback echt is. Moffett (1968) is een voorstander van leerlingenrespons, omdat responses echt behoren te zijn, passend bij de actie. Hij is een tegenstander van de gestandaardiseerde, professionele reactie. Ook Elbow (1974) pleit voor respons door leerlingen wegens de authenticiteit van dergelijke feedback. De inhoud van de feedback van leerlingen zou volgens didactici de volgende kenmerken vertonen: de boodschap is constructiever dan het docentencommentaar; draagt meer betekenis, omdat de boodschap in leerlingentaal en niet in vaktaal is geschreven; is behulpzamer; is gedetailleerder omdat leerlingenlezers meer tijd besteden aan een opstel dan docenten; is uitgebreider; is levendiger, is eerlijker, direct op de man af.

Al deze kwaliteiten dragen ertoe bij dat schrijvers meer met de feedback gaan doen, omdat schrijvers meer geïnteresseerd zijn in de feedback van hun klasgenoten, meer waarde hechten aan het leerlingencommentaar, ontvankelijker zijn voor de feedback omdat de feedbacksituatie als informeler wordt ervaren en als minder bedreigend, wellicht door het ontbreken van een cijfer. Ze willen het commentaar echt proberen te begrijpen. De positieve houding jegens de feedbackboodschap kan ook veroorzaakt worden door het enthousiasme waarmee de responstaak werd uitgevoerd. De zorg die schrijvers zelf besteedden aan de uitvoering van de feedbacktaak toen zij op andermans teksten moesten reageren, kan ertoe bijdragen dat zij vertrouwen stellen in de respons die zij zelf krijgen.

Voor het herschrijven betekent dit dat leerlingen het herschrijven niet langer zien als een nodeloze tijdsinvestering, als een strafmaatregel, maar als een uitdaging om tegemoet te komen aan de commentaren. De eisen die leerlingen stellen zijn haalbaarder dan de normen van de docent, die te sophisticated tegenover leerlingenprodukten staat. Te verwachten valt dat schrijvers intensief en grondig zullen reviseren in het geval

van leerlingencommentaar. Het zou kunnen zijn dat vooral minder goede leerlingen baat hebben bij leerlingenrespons, omdat juist zij het lerarenoordeel vrezen.

Moffett (1968) acht de egocentrische instelling van de schrijvers hèt probleem waar het schrijfonderwijs mee kampt. Juist leerlingenrespons stelt dit probleem aan de orde. Veelal denken leerlingen dat de lezers precies hetzelfde denken als zijzelf toen ze de tekst schreven en dat lezers alle impliciete denkstappen die de schrijvers vanuit hun referentiekader kunnen maken, ook zullen nemen. Wat leerlingen nu vooral leren door de respons op hun opstellen is, dat er voor iemand geschreven wordt, dat de handeling ook een indirect object heeft, een geadresseerde. Zij gaan beseffen dat zij in hun opstel hún waarneming van de werkelijkheid in een tekst weergeven en dat zij veelal verzuimen rekening te houden met het inbedden van de ideeën in een context, het op het spoor zetten van de lezers opdat zij zullen lezen wat de schrijver bedoelt. Door de interactie over teksten, stellingen en ideeën, zijn leerlingen in de gelegenheid vanuit lezersperspectief naar de werkelijkheid te kijken. Ze worden zich bewust dat het van groot belang is zich heel duidelijk voor de geest te halen hoe zijzelf de wereld waarnemen, en dat het daarnaast belangrijk is na te gaan hoe lezers de wereld en de ideeën waarnemen zoals ze in de tekst verwoord worden. Helderheid wordt een belangrijke communicatie-eis, omdat schrijvers toch willen dat lezers lezen wat zijzelf bedoelden te zeggen. De zorg om helderheid en grammaticaliteit blijkt geen eigenaardige professionele zorg te zijn van de docent, maar een reële zorg van reële lezers

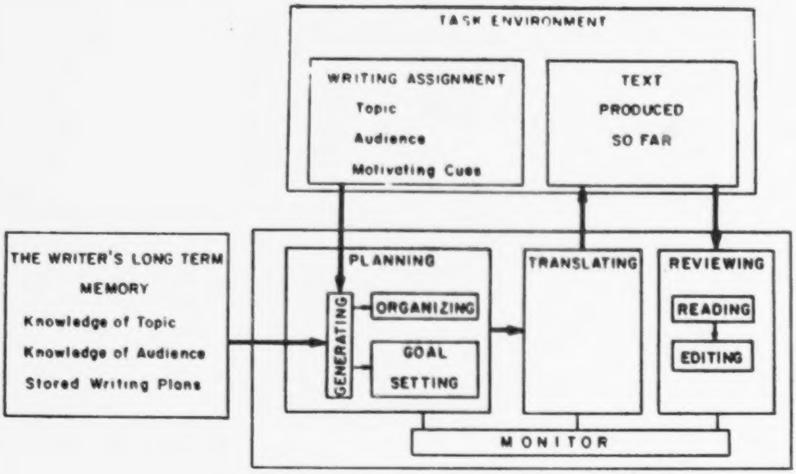
Deze zorg voor helderheid, explicitering, contextverzorging om de ruis in de communicatie te verminderen, wordt nog eens versterkt doordat het schrijvers blijkt, dat er veel manieren zijn waarop verscheidene lezers op één en dezelfde tekst reageren. Naast de criteria van helderheid en duidelijkheid die schrijvers afleiden uit de respons, bevat de responsboodschap meer criteria over een goede tekst. Criteria zijn ook in het geding, als leerlingen respons krijgen waarin suggesties gedaan worden of vragen gesteld worden.

## 2.3. Schrijfprocestheorie

Hayes en Flower (1980; 1983) ontwikkelden op basis van hardop schrijfprotocollen en in aansluiting op probleemoplossingsmodellen een model van het schrijfproces. Het model sluit aan bij wat al intuïtief door docenten aan processen werd onderscheiden: plannen, organiseren, reviseren zijn oude bekenden. Het nieuwe van het procesmodel is echter, dat de organisatie van het stelproces beschreven wordt als een doelgericht, recursief probleemoplossingsproces. Na een korte beschrijving van het model, gebruikmakend van Hayes en Flower (1980, 1983), Flower en Hayes (1981) en Bochardt (1983), worden de effecten van leerlingenrespons in dit theoretische kader geïnterpreteerd.

Het model kent drie componenten (zie figuur 2.2. hieronder): de taakomgeving, het lange termijngeheugen en het schrijfproces. De taakomgeving staat voor de verzameling van alle externe stimuli die de uitvoering van het schrijfproces beïnvloeden: de schrijfopdracht (waarin opgenomen onderwerp, publiek en factoren die de motivatie bevorderen) en de tekst voor zover die af is, omdat de tekst ook een repeterende invoer is voor het schrijfproces. In het lange termijngeheugen bevindt zich alle kennis die de schrijver gebruikt bij het concipiëren van tekst als kennis van de wereld (van het onderwerp), kennis van lezers, kennis van tekstschema's.

Figuur 2.2.: Het schrijfprocesmodel van Flower en Hayes




Uit: Gregg en Steinberg (1980)

In het schrijfproces onderscheiden Flower en Hayes drie subprocessen: plannen, formuleren en teruglezen. In het planningsproces roept de schrijver de informatie-elementen op die mogelijk in de tekst gebruikt zullen worden (genereren), groepeert en selecteert de ideeën (organiseren) en stelt zich doelen (doelbepalen) tijdens het uitvoeren van de taak: de schrijver stuurt het proces ('Hier moet nog een voorbeeld', of 'Laat ik eerst maar opschrijven wat me te binnen schiet'). Het formuleerproces zet de ideeën om in taal. Het terugleesproces heeft tot doel de tekst te verbeteren: de schrijver leest passages, zinnen, woorden over, om na te gaan wat moet volgen en om te controleren of wat er staat strookt met intenties van de schrijver, het effect op de lezer en de eisen die men aan het taalgebruik stelt. De processen worden gestuurd door een monitor, die tevens de uitvoer van de processen controleert.

Hoe ziet de taakomgeving eruit als een leerling voor een publiek van medeleerlingen schrijft en weet dat die medeleerlingen de tekst zullen lezen en erop zullen reageren? Volgens de practici ervaren leerlingen de schrijfopdrachtsituatie als een communicatieve situatie: er is nu in de taakomgeving een extra externe stimulus: de toekomstige lezer. Daardoor zou het in de opdracht opgenomen taakelement 'audience' een sterkere signaalfunctie kunnen krijgen. De wetenschap dat medeleerlingen de tekst zullen gaan lezen werkt volgens de practici als een 'motivating cue'. Omdat de afstand tussen schrijver en lezer klein is, zal een schrijver gemakkelijker aan de slag gaan.

Als hij begint, zal hij over meer kennis in het lange termijn geheugen kunnen beschikken, dan wanneer hij nog niet in een proces van leerlingenrespons verwikkeld was geweest. Door het lezen van andermans opstellen heeft de schrijver de eigen prestatie met die van anderen kunnen vergelijken. Hij ziet dat er veel (goede) uitwerkingen kunnen zijn van één opdracht. Er wordt een voorraad modellen van teksten, inleidingen, stijlen, en retorische middelen opgebouwd. Hij weet meer over wat er mis kan gaan in communicatie via teksten. Hij kent de lezers via hun opstellen. Door het vervullen van de commentaartaak ontwikkelt de schrijver kennis van criteria voor goede teksten. Die kennis wordt uitgebreid door de respons die hij op de eigen opstellen kreeg.

De kenmerken van de taakomgeving en de inhoud van het geheugen hebben invloed op het verloop van de denk- en schrijfprocessen. Zo zullen in het genereerproces meer criteria worden gegenereerd ten behoeve van het redigeerproces. Het organiseerproces zal gebruik kunnen maken van modellen uit het geheugen. Omdat er meer zicht is op het gewenste eindprodukt, kan het gehele schrijfproces doelmatiger en doelgerichter verlopen. Het formuleerproces zal bij leerlingenrespons in het teken van het criterium duidelijkheid staan. Leerlingen hebben als lezer en responsontvanger ervaren dat een tekst niet hoeft te zijn wat de schrijver bedoelde en zullen trachten de slaagkans van de communicatie met de lezer te verhogen. Verwacht wordt dat het formuleerproces onderbroken wordt om formuleringen af te wegen op duidelijkheid en effect.

Het redigeerproces zal onder invloed van de gegenereerde criteria zorgvuldig gestuurd worden. De zelfkritische houding groeit; het op afstand waarnemen van de tekst wordt geleerd.

Tot wat voor tekstkwaliteiten zou leerlingenrespons leiden? Het is alle steldidactici er uiteindelijk om te doen om hun leerlingen betere teksten te laten produceren. Veel auteurs die leerlingenrespons voorstaan schrijven dan ook dat de teksten die leerlingen schrijven beter worden. Er zijn uitspraken over de algehele verbetering van teksten. Volgens Hoover (1972) leren vooral minder goede leerlingen betere teksten te schrijven en consolideren goede leerlingen hun vaardigheid. Meer specifieke kenmerken waarop de teksten vooruitgaan, zijn volgens de practici uit par. 2.2. de volgende.

De inhoud van teksten wordt helderder, concreter, rijker en completer; minder generaliserend, meer gedetailleerd en onderbouwd. De inhoudselementen zijn alle gekoppeld aan een basiszin. Organisatiefouten worden nauwelijks meer gemaakt. De opbouw van de hele tekst wordt logischer. Er wordt vaker gebruik gemaakt van een basisstructuur. De stijl wordt individueler, levendiger en vloeiender. De teksten geven blijk van publiekgericht schrijven: er worden meer lezersgerichte zinnen in de teksten opgenomen, meer introducerende zinnen en meer verbindende zinnen. Het aantal fouten dat het begrip van de lezer bemoeilijkt neemt af, evenals fouten op het gebied van conventies (spelling, interpunctie, grammatica en stijl). De teksten ogen netter en worden langer.

In diverse onderzoeken is getracht verschillen vast te stellen in de schrijfprocessen van goede en slechte schrijvers. Overzichten over deze onderzoeken (Bochardt 1983, 1984; Humes, 1983; Scardamalia en Bereiter, 1986) presenteren globaal eenzelfde beeld van deze verschillen. Goede schrijvers hebben meer macht over hun tekst: zij plannen meer, sturen het proces, beschouwen de tekst van een afstand en reviseren grondig. Slechte schrijvers zijn in de ban van de zich ontwikkelende tekst en bezien de tekst alleen op het voorkomen van fouten.

## 2.4. Onderwijskundige theorieën

Leerlingen die elkaars teksten becommentariëren verrichten een instructietaak die traditioneel door de docent wordt uitgevoerd. Als leerlingen door een dergelijke onderwijsleersituatie betere prestaties leveren, moeten daar onderwijskundige verklaringen voor zijn. In een overzicht van onderzoeken over 'peer teaching' en 'tutoring' in leesonderwijs toont Bloom (1976) aan dat zowel de onderwijsontvangenden als onderwijsgevenden van de onderwijsleersituatie profiteren. Het leereffect voor de onderwijsontvangenden verklaart zij met behulp van drie leerprincipes, afgeleid van Dollard en Miller (1950). In een één-op-één-relatie zijn de aanwijzingen ('cues') duidelijker doordat in de interactie met de leerling de lesgever de instructie steeds kan verhelderen en aanpassen als blijkt dat dat nodig is. De activiteit ('participation') van de leerling is groot: de mogelijkheid om actief te reageren is veel groter dan in groepsinstructie. De lesgever kan de reacties meteen beoordelen, zonodig corrigeren of bevestigen. In individuele instructie is er meer kans op adequate bemoediging ('reinforcement'): de onderwijsleersituatie is heel privé, waardoor de angst van de leerling afneemt. De leerling zal daardoor vrijer deelnemen aan het onderwijs, waardoor de lesgever meer kans heeft feedback te geven. Binnen een korte tijd leert de lesgever die reacties van bemoediging te geven, die effectief blijken voor juist die leerling.

Bloom (1976) maakt duidelijk dat niet alleen de onderwijsontvangenden maar ook de onderwijsgevenden baat hebben bij 'peer teaching' of 'tutoring': de resultaten in het onderwezen vak verbeteren, alsmede de attitude jegens schoolgaan en/of het zelfbeeld. Vooral als de pupil van de onderwijsgevende leerling leerwinst boekt, stijgt het resultaat van de onderwijsgevende leerling. Bloom brengt, merkwaardigerwijs, het leren van lesgevers niet in verband met de drie hiervoor genoemde leerprincipes. Zij veronderstelt dat lesgevers de lesstof intensiever bestuderen, vooral datgene waarop zij nog niet voldoende greep hadden. Zij worden daartoe gemotiveerd doordat zij zich moeten bewijzen ten overstaan van een echt publiek. Terwijl zij zich voorbereiden de lesstof interessant te maken voor hun leerling, bewerken zij die stof voor zichzelf. De affectieve groei zou bevorderd worden door het dragen van verantwoording en de voldoening van het succes van hun leerling.

Voor de verklaring van de leerresultaten op het gebied van stelvaardigheid zijn de drie principes die Bloom noemt, 'cues', 'participation' en 'reinforcement', niet adequaat als we de onderwijssituatie van 'leerling-leerlingonderwijs' vergelijken met docent-leerlinginstructie. De docent die een opstel van een leerling becommentarieert, verkeert ook in een één-op-één-situatie en zal de 'reinforcement' kunnen afstemmen op de individuele leerling. Zijn commentaar zal meteen individueel aangepaste aanwijzingen ('cues') bevatten voor een volgende schrijftaak. De participatie van een leerling in traditioneel stelonderwijs zal niet minder zijn als er evenveel schrijfprodukten verlangd worden als in de situatie waarin een leerling zijn opstellen becommentarieert. Echter, in beide onderwijsleersituaties zal niet zoveel geschreven worden, dat lesgevers veelvuldig feedback kunnen geven en kunnen bemoedigen. De situatie verschilt sterk van die in leesonderwijs, waarop Bloom haar analyses steunt, waarin de lesgever op veel meer momenten corrigerend en bemoedigend kan optreden.

Ter verklaring van leereffecten van lesgevers in stelonderwijs kunnen de drie principes echter wel gehanteerd worden. Door het vervullen van de feedbacktaak ontwikkelen de leerlingen zelf meer duidelijkheid over de 'cues': 'the stimuli which convey to students what they are to do and how they are to do it' (Bloom 1976, p.5). Zij zijn actief in de onderwijsleersituatie betrokken ('participation'). Het becommentariëren en het leren daardoor kan fungeren als 'reinforcement', zeker als de ontvangende leerling er blijk van geeft het commentaar te waarderen.

Blooms principes gelden dus vooral als een vergelijking wordt gemaakt tussen groepsinstructie en individuele instructie. Stelonderwijs is in feite al sterk geïndividualiseerd onderwijs, aangezien het accent van de instructie op de bespreking van het individuele resultaat ligt. Daarom voldoen de principes niet om de effecten van leerlingenrespons in stelvaardigheidsonderwijs te verklaren als het gaat om de effecten voor de leerling in de rol van leerder. Wel zijn zij geschikt om de leereffecten te verklaren van de leerling in zijn rol van onderwijsgevende.

Ten Cate, Tromp en Cornwall (1984) besteden in hun boekje over de student als docent ruim aandacht aan mogelijke verklaringen voor leereffecten. Zij maken onderscheid tussen 'leren door doceren' en 'leren door onderwijs van medeleerlingen', de twee rollen die wij in Figuur 2.1. onderscheidden als componenten van het instructieve koppel.

De roltheorie van Sarbin (1976; Sarbin en Allen, 1968) wordt als motivationele verklaring aangedragen. Sarbin onderscheidt rollen met een verschillende status (verkregen versus toegeschreven status) waaraan verschillen in beloningsvorm en betrokkenheid gekoppeld zijn. Een leraar kan zich losmaken van zijn rol; hij is leraar geworden uit eigen keuze (verkregen status). Leerling wordt men automatisch (toegeschreven rol). Een leerling kan zich nauwelijks uit zijn rol losmaken. De beloning voor een verkregen rol geschiedt in symbolen van achting (salaris, cijfers), voor de toegeschreven rol zijn uitingen van zorg beloningen. In het onderwijs zijn de relatie en de beloningsvormen echter vooral gebaseerd op een verworven status voor zowel de docent (terecht) als de leerling (niet terecht). Cijfers als beloning zijn niet voldoende: een leerling heeft wegens zijn rolstatus ook uitingen van zorg nodig. Deze uitingen worden minder frequent naarmate de leerkracht van pedagoog in vakleerkracht verandert, ofwel naarmate men op een hoger niveau onderwijs volgt. Als leerlingen echter elkaar onderwijzen zullen zij wel effectieve, bij de toegeschreven rol passende beloningswijzen hanteren. Een leerling gaat door de zorg en vriendschap in die relatie harder werken.

Door het vervullen van de rol van docent, ervaren leerlingen gevoelens en vertonen zij gedragingen die met een dergelijke rol in overeenstemming zijn: prestige, autoriteit en competentie. Zeker als de taak met enig succes wordt verricht, geeft de rolvervulling aanleiding tot het opbouwen van een positief zelfconcept. Ten Cate, Tromp en Cornwall (1984) halen theorieën van Bandura (1982) en Weiner (1974) aan om die ontwikkeling te verklaren.

De verklaringen van Ten Cate et al. ondersteunen Blooms (1976) veronderstellingen die hiervoor beschreven werden: de wijze waarop een leerling een andere leerling bemoedigt, is van een andere aard dan die waarop de docent een leerling bemoedigt. Het vervullen van de docentrol schept een potentiële mogelijkheid voor het ontwikkelen van een positief zelfconcept.

Behalve een motivationele verklaring poneren Ten Cate et al. een cognitieve verklaring voor het effect van 'leren van medeleerlingen' en 'leren door te doceren': de theorie van het semantisch netwerk. Een leerling die iets bestudeert omdat hij het anderen moet onderwijzen zal een ander semantisch netwerk opbouwen van de leerstof dan wanneer hij de stof in de rol van leerling als leertaak opgedragen krijgt. Aan de andere kant: leerlingen die elkaar onderwijzen verschillen qua semantisch netwerk minder van elkaar dan een docent en een leerling. De congruentie tussen de netwerken maakt dat de communicatie

tussen de participanten op eigen niveau plaatsvindt. In feite is er sprake van een natuurlijke aansluiting bij het kennis- en taalniveau van de leerling. Toegepast op leerlingenrespons in stelonderwijs lijken de roltheorie en de informatieverwerkingstheorie houdbaar. De onderwijsleersituatie wordt als veiliger ervaren. De feedbackgevers zijn 'peers', significante anderen, wier commentaren anders gewaardeerd worden dan die van docenten. Het vervullen van de commentatorrol geeft trots en respect. Naast die motivationele effecten léért de leerling anders, omdat hij criteria voor goede teksten zelf moet toepassen en deze niet alleen als respons ontvangt. Het commentaar dat hij ontvangt sluit beter aan bij zijn semantische netwerk omdat het lezersrespons is en geen vakman-commentaar.

## 2.5. Conclusie en hypothesen

Via een analyse van uitspraken van practici over effecten van leerlingenrespons en oorzaken daarvan, kwamen wij tot een beschrijving van relaties tussen kenmerken van de onderwijsleersituatie en effecten op stelvaardigheid. Daarin werden vier fasen, - schrijven, lezen, commentaar geven, commentaar ontvangen -, verbonden met gerapporteerde effecten. Deze waren van motivationele en vakinhoudelijke aard. De gerapporteerde effecten lijken te passen in onderwijskundige theorieën die als verklaringen voor effecten van leerling-leerlingonderwijs geponeerd zijn. De effecten lijken eveneens verklaarbaar vanuit de theorie die Hayes en Flower (1980; 1983) ontwikkelden over het schrijfproces.

Van het inschakelen van leerlingen als lezer en commentator van teksten van medeleerlingen mag op goede gronden effect verwacht worden. Effecten verwachten we op drie terreinen: schrijfprestaties, attituden en zelfconcept, schrijfprocessen.

Het inschakelen van leerlingen als commentatoren van elkaars teksten zou ertoe leiden dat de teksten meer tekenen zouden gaan vertonen van het besef dat schrijven communiceren is: de teksten worden doel- en publiekgericht. Zij zijn niet een weerspiegeling van de inhoud zoals die in het geheugen van de schrijver opgeslagen ligt, maar een bewerking daarvan (vgl. Flowers (1979) begrippen 'writer based prose' en 'reader based prose'). Tengevolge van de communicatieve situatie en de verworven modellen zullen opstellen beter worden qua stijl en opbouw.

Verwacht wordt verder dat leerlingen door leerlingenrespons meer vertrouwen in het eigen kunnen krijgen. Zij krijgen meer plezier in het schrijven. Het laten lezen of laten beoordelen van het opstel zullen leerlingen minder bedreigend gaan vinden. De houding jegens het beoordeeld worden, wat elke schriftelijke communicatieve handeling impliceert, wordt positiever.

Schrijfprocessen zullen door leerlingenrespons veranderen. Door het lezen van veel andere opstellen zal de voorraad inhouden, presentatiestructuren en formuleringen toenemen. Leer-

lingen zullen tijdens het proces meer alternatieven genereren wat inhouden, presentatiestructuren en formuleringen betreft. Door het uitvoeren van de commentaartaak en de aanwezigheid van een reëel publiek, is te verwachten dat leerlingen meer criteria genereren voor een goede tekst. Deze criteria zullen bij het revisieproces een rol spelen. Er zal meer gereflecteerd worden op de ontwikkelde tekst vanuit een lezersperspectief. Het schrijven van een 'reader based' tekst veroorzaakt een intensiever ordenen van gegenereerde informatie. Het valt te verwachten dat leerlingen hun schrijfproces meer sturen omdat zij beter weten waarheen het moet leiden: het zal meer gepaard gaan met zelfinstruerende opmerkingen. Hun revisieproces zal op dat van goede schrijvers gaan lijken: ze zullen het aandurven om niet alleen op lokaal niveau te verbeteren, maar ook op tekstniveau.

## 3. ONDERZOEKSLITERATUUR

Voordat een experimentele stelcursus ontworpen werd met leerlingenrespons als kenmerk van de feedbackfase, werd onderzoeksliteratuur bestudeerd. Doel daarvan was inzichten te verwerven in produktieve en contraproduktieve factoren in het onderwijsleerproces waarin leerlingen elkaars werk beoordelen. Verder zou onderzoeksliteratuur waarin experimenten uitgevoerd zijn die te vergelijken zijn met het onze ons kunnen wijzen op problemen in het opzetten en uitvoeren van dergelijke experimenten. Resultaten van dergelijke onderzoeken konden wellicht een basis verschaffen voor een bijstelling van de keuze van afhankelijke variabelen.

In paragraaf 3.1. worden een klein aantal studies beschouwd waarin storende variabelen aangetoond zijn. Dit zijn studies waarin het algemene proces van 'peer evaluation' thema is. Vervolgens komen in par. 3.2. studies aan bod waarin het proces van leerlingenrespons bij stelonderwijs enigszins uitgewerkt is. In par. 3.3. worden 21 quasi-experimentele effectstudies op vier aspecten geanalyseerd. Conclusies trekken we in par. 3.4.

### 3.1. Storende invloeden bij collega-beoordelingsprocessen

Wie een stelcursus ontwerpt waarin leerlingenrespons een belangrijk onderdeel is, doet er goed aan rekening te houden met beschikbare kennis over variabelen die het beoordelingsproces verstoren. Over variabelen die een ongewenste invloed uitoefenen op beoordelaars is veel literatuur. Voor deze algemene paragraaf hebben we ons echter beperkt tot literatuur over de specifieke beoordelingssituatie waarin de beoordelaar een gelijke is van de beoordeelde: een medeleerling, een medestudent, een collega. De variabelen die we aantroffen zijn contaminatie- en halo-effecten: beoordelaars gebruiken - bewust of onbewust - de beoordeling voor andere doeleinden dan bedoeld is, respectievelijk laten hun oordelen beïnvloeden door niet relevante kenmerken van het te beoordelen produkt. (De Groot, 1972). De variabelen die hierna aan de orde komen zijn vriendschapsrelaties, sexe en een variabele die omschreven kan worden als 'image-building': als een beoordelaar zich op zijn beurt beoordeeld weet, dan verstoort dat zijn beoordeling. Deze drie storingsbronnen gelden niet alleen de beoordelaars, zij blijken ook het accepteren van de feedbackboodschap te beïnvloeden. Bevindingen van een aantal studies zullen hierna in deze volgorde besproken worden. Vervolgens wordt ingegaan op mogelijke effecten van twee kenmerken van de situatie waarin collega's oordelen van collega's krijgen: de openbaarheid en de ambiguïteit van de oordelen. De resultaten van deze studies zullen leiden tot overwegingen bij de constructie van het experimentele onderwijsprogramma. Die volgen aan het eind van dit hoofdstuk.

...

Leerlingen in een klas zijn via allerlei affectieve relaties met elkaar verbonden. Leerlingen kennen elkaar goed doordat er veel interactie is binnen en buiten schooltijd. Zij kennen elkaar vaak al langer dan één schooljaar. Te verwachten is dat relaties tussen leerlingen elkaars beoordelingen beïnvloeden. Schrader en Westphal (1976) verrichtten onderzoek hiernaar. In hun onderzoek bleek dat leerlingen vriendjes een hoger cijfer gaven als zij medeleerlingen moesten beoordelen in practica over verkoopgesprekken. Dit effect nam iets af als in plaats van een algemene indruk een cijfer gegeven moest worden op basis van tevoren vastgestelde criteria: het gemiddelde cijfer lag dan iets lager en de scores vertoonden een grotere spreiding.

Cederblom en Lounsbury (1980) onderzochten welke weerstanden bij beoordeelden samenhingen met de mate waarin zij deze collega-beoordeling accepteerden. Vriendschapsrelaties en (on)bruikbaarheid van de feedback bleken de sterkst ervaren weerstanden. Volgens de respondenten beïnvloedden vriendschapsrelaties de nauwkeurigheid van beoordelingen sterk. De onderzoekers bevelen aan criteria te specificeren waarop feedback gegeven moet worden; dit zal de feedbackwaarde voor de beoordeelden verhogen en de waargenomen verstoringen door vriendschapsrelaties doen afnemen (vgl. Schrader en Westphal, 1976). Een tweede aanbeveling betreft het diagnostisch dan wel summatief gebruik van collega-beoordeling. Volgens Cederblom en Lounsbury zal de feedbackwaarde toenemen als collega-beoordeling gebruikt wordt als 'tussenstand': de beoordeelden moeten in de gelegenheid gesteld worden iets met de feedback te doen en daar eventueel voor beloond worden.

Uit feedbackliteratuur over docenten-leerlingen interactie, blijkt dat met sexe van docenten en leerlingen verschillen samenhangen in de hoeveelheid en de aard van de feedback die docenten geven (Dweck et al.,1978). Dit verschijnsel blijkt zich ook voor te doen als leerlingen elkaar beoordelen. Soares en Soares (1980) stelden in hun onderzoek vast dat jongens èn meisjes aan jongens hogere scores toekenden dan aan meisjes. Book en Wynkoop Simmons (1980) onderzochten de aard van de feedback in zulke situaties. Zij analyseerden daartoe het commentaar dat leerlingen elkaar gaven op hun spreekbeurten. Het patroon van commentaar dat meisjes geven en krijgen was: op presentatie gerichte, negatieve, persoonlijke en atomistische feedback. Jongens gaven daarentegen juist op inhoud gerichte, positieve, onpersoonlijke, holistische feedback. De feedback die zij krijgen komt hiermee overeen; alleen is deze atomistisch i.p.v. holistisch. Book en Wynkoop Simmons (1980) troffen daarnaast nog een interactie aan: meisjes krijgen van sexegenoten minder positief commentaar dan zij van jongens krijgen.

Het onderzoek van Dweck et al. (1978) gaat in op de effecten van de feedback bij jongens en meisjes: zij blijken verschillend te reageren op de persoon van de feedbackgever. De onderzoekers gingen na waaraan jongens en meisjes hun falen

toeschrijven. Als de feedback door de docent gegeven wordt, zijn meisjes geneigd hun falen toe te schrijven aan een gebrek aan vaardigheid. Hun verwachtingen om te zullen slagen in een volgende taak zijn dan niet hooggespannen, omdat een stabiele trek, - een gebrek aan vaardigheid -, niet makkelijk te veranderen is. Is de docent een vrouw, dan treedt dit effect nog sterker op. Als de feedback door leerlingen gegeven wordt, dan zijn het de jongens bij wie dit attributieproces zich voordoet. Dit zou kunnen betekenen dat docentfeedback voor falende meisjes en leerlingenfeedback voor falende jongens funest is: de moed ontbreekt hun om zich te motiveren voor een volgende taak. Ze aanvaarden hun falen op dit soort taken als hun noodlot.

Of dit proces van attribueren zich ook voordoet als leerlingen, zoals in een steldidactiek van leerlingenrespons, in de gelegenheid zijn hun prestatie te verbeteren door hun tekst te herschrijven, is niet zeker. Door die gelegenheid te scheppen zouden leerlingen er wellicht meer toe neigen het resultaat van de eerste versie toe te schrijven aan inzet, een wel veranderbare factor. Een dergelijk proces kan wel aanleiding zijn om de effecten van zo'n steldidactiek niet alleen voor jongens en meisjes apart te onderzoeken, maar ook binnen die twee groepen te differentiëren naar beginbekwaamheid.

Beoordelaars blijken gevoelig voor degenen die hun beoordelingen waarnemen. Amabile en Glazenbrook (1982) verrichtten twee experimenten om het effect te onderzoeken dat zij omschrijven als de tendens anderen te deklasseren uit oogpunt van zelfhandhaving, zelfs al is er geen sprake van een negatieve relatie tussen de betrokkenen. Het bleek dat leerlingen elkaar negatievere beoordelingen gaven dan nodig was in twee situaties: als zij dachten dat hun oordelen door een beoordelaar met een hoge status (een professor bijvoorbeeld) werden waargenomen of als zij binnen een groep leerlingen in het onzekere verkeerden over hun positie binnen die groep. Via de beoordelingen moesten zij bewijzen dat zij beter waren dan de andere leerlingen.

Bradley (1981) onderzocht wat artikelenschrijvers van het commentaar vonden dat ze kregen van medewerkers aan tijdschriften. Een opvallend resultaat is dat beoordeelden hier veel hinder zeiden te ondervinden van het 'image-building'-effect dat Amabile en Glazenbrook (1982) constateerden: schrijvers hadden sterk het idee dat beoordelaars van manuscripten zochten naar tegenwerpingen om de redacteur te tonen dat ze goed hun best hadden gedaan. Dit effect beïnvloedt het accepteren van de feedback in negatieve zin. De respondenten waren ook ontevreden over de persoonlijke voorkeuren van beoordelaars, de objectiviteit en de competentie: vaak was een beoordelaar niet goed thuis op het terrein dat het artikel bestreek. Deze mankementen aan het beoordelingssysteem klemden te meer, omdat men verplicht werd te reviseren op voorschriften van de redacteur die gebruik maakte van de beoordelingen.

Een verschil tussen het beoordelen door een docent en door leerlingen, is de mate van openbaarheid: als leerlingen beoordelen dan nemen zij kennis van een aantal produkten van medeleerlingen. Er zijn meer mensen die weet hebben van de prestatie van een leerling. Welke effecten heeft een dergelijke beoordelingssituatie op de beoordeelde leerling? Er zijn vier onderzoeken die ingaan op deze effecten. Uit de studies van House (1980) en Salili et al. (1976) blijkt dat de motivatie van beoordeelden aangemoedigd wordt. White (1972) toont aan dat leerlingen in die situatie minder geneigd zijn zichzelf verantwoordelijk te stellen voor het resultaat. De onderzoeksresultaten van Silver en Schlenker (1981) wijzen uit dat leerlingen zich sterker zullen conformeren aan de oordelen van medeleerlingen als die ook gekend worden door anderen, zoals de onderzoeker of de docent. Deze vier onderzoeken zullen hieronder kort behandeld worden.

In een experiment van House (1980) bleken leerlingen in de situatie waarin andere leerlingen kennis hadden van hun prestatie, falen niet zo vaak toe schrijven aan een gebrek aan vaardigheid als aan een gebrek aan inzet. Is er sprake van slagen in een openbare beoordelingssituatie, dan schrijven beoordeelde leerlingen dat toe aan hun vaardigheid. In het kader van Weiners motivatietheorie (1972) zijn beide processen als positief te waarderen. Leerlingen die een goed resultaat toeschrijven aan hun vaardigheid, zullen met vertrouwen een nieuwe taak aanpakken, omdat vaardigheid een stabiele factor is die min of meer garant staat voor nieuw succes. Zij beginnen met een positieve 'Affective Entry' (Bloom, 1976). Leerlingen die faalden, pakken een volgende taak met meer inzet aan. Openbaarheid van de beoordelingssituatie zoals bij leerlingenrespons zou dus positief op de motivatie kunnen werken.

De resultaten van het onderzoek van Salili et al. (1976) ondersteunen deze veronderstelling. Zij lieten leerlingen taken uitvoeren in drie verschillende beoordelingscondities. Deze verschilden in de persoon die de kwaliteit van de prestatie te weten zou komen: niemand anders dan de leerling zelf; de medeleerlingen of de docent. Deze drie condities hadden geen effect op de prestatie, wel echter op wat de onderzoekers 'Continuing Motivation' noemen. De wens om door te gaan met dergelijke taken was in de conditie waarin de docent het resultaat van de prestatie te horen zou krijgen het zwakst; tussen beide andere condities werd geen verschil in motivatie geconstateerd: leerlingen waren in beide condities intrinsiek gemotiveerd.

White (1972) onderzocht drie soortgelijke condities op de mate waarin leerlingen zichzelf verantwoordelijk stellen voor het resultaat van hun prestatie. Het gaat hier dus ook om wat House (1980) interne attribuering noemde: vaardigheid en inzet. Het bleek dat leerlingen die hun werkstukken zelf beoordeelden het meest zichzelf verantwoordelijk stelden voor de kwaliteit van het produkt. Leerlingen, wier werkstukken door 14 medeleerlingen in het openbaar beoordeeld werden,

vertoonden die zelfverantwoordelijkheid het minst. White wijt het falen van het beoordelen door medeleerlingen op dit punt o.a. aan de grote groep beoordelaars: individuele leerlingen hadden te weinig kans om invloed uit te oefenen op de normen waarmee het werk werd beoordeeld. Als beoordelaar was zijn stem slechts een van de veertien en als zijn werk beoordeeld werd, stond hij tegenover een overmacht. Leerlingen moesten zich dus conformeren aan de groep, waardoor er weinig gelegenheid was een eigen standaard te ontwikkelen waaraan leerlingen de eigen werkstukken konden beoordelen. De normen komen te zeer van buitenaf, zijn te zeer extern, zodat leerlingen zichzelf niet verantwoordelijk stellen voor hun resultaten. Leerlingen zijn niet in de gelegenheid een interne 'locus of control' te ontwikkelen.

De onderzoeksresultaten van Silver en Schlenker (1981) sluiten in zekere zin aan op het bovenstaande. Deze onderzoekers gingen onder meer na wat het effect is van het al dan niet openbaar zijn van beoordelingen door medeleerlingen. Daartoe verstrekten zij leerlingen gefingeerde, negatieve testresultaten en lieten zij leerlingen voor zichzelf of in het bijzijn van anderen een zelfbeoordeling maken. Als leerlingen nu voor zichzelf reageerden op de testresultaten werden hun zelfbeoordelingen niet beïnvloed door de testresultaten. Reageerden zij daarentegen in het bijzijn van anderen, dan trokken de zelfbeoordelingen wèl sterk naar de onjuiste testresultaten. Dus als testresultaten openbaar zijn, voelen leerlingen zich verplicht zich te conformeren aan die testresultaten, hoe onterecht dat ook is. Er zal in zo'n situatie geen sprake kunnen zijn van het opbouwen van een zelfverantwoordelijkheidsgevoel, omdat de leerlingen onder druk van externe omstandigheden reageren.

In de literatuur over het beoordelen van produktieve taalvaardigheden (zie voor een overzicht Wesdorp, 1981) blijkt steeds weer dat oordelen van docenten en leerlingen onbetrouwbaar zijn, dat wil zeggen dat beoordelaars het onderling niet erg eens zijn over één en dezelfde prestatie. Bij leerlingenrespons mag men dan ook verwachten dat leerlingen het onderling niet al te zeer eens zullen zijn over de kwaliteiten van een tekst die een medeleerling schreef. Sommige didactici, met Elbow (1974) als voorman, vinden die verscheidenheid aan oordelen juist een groot goed van leerlingenrespons: schrijvers ervaren dan wat er bij verschillende lezers gebeurt als ze één en dezelfde tekst lezen en kunnen daar bij het herschrijven hun voordeel mee doen. De vraag is echter wel, wat leerlingen nu precies doen met ambigue oordelen. Sherwood (1967) ging na wat het effect is van ambigue oordelen op de beoordeelde. Hij veronderstelde dat leerlingen bij eensluidende oordelen van medeleerlingen, deze oordelen zouden volgen: de correlatie tussen eensluidende beoordelingen door andere leerlingen en de zelfbeoordeling die daarop volgde zou hoog zijn. Sherwood trof inderdaad een hoge correlatie aan tussen oordelen van anderen met weinig variantie en zelfbe-

oordelingen. Het eigen oordeel lag ook dicht bij het gemiddelde van oordelen die anderen gaven. Als de oordelen van anderen echter veel variantie vertoonden, beoordeelden leerlingen zichzelf steevast hoger dan het gemiddelde van de andere oordelen. De beoordeelde leerlingen bleken in hun oordeel het meest beïnvloed door het oordeel van vrienden.

In deze paragraaf zijn enkele resultaten van onderzoeken over collega-beoordeling bijeengebracht, die op vijf factoren wijzen die invloed uitoefenen op het geven en het ontvangen van collega-beoordelingen: affectieve relaties tussen de betrokkenen, sexe van betrokkenen, de wetenschap van beoordelaars dat hun beoordelingen beoordeeld worden, het meer openbare karakter van de beoordeling en ambiguïteit in oordelen. In par. 3.4., Conclusies, komen we op de gevolgen van deze factoren voor de constructie van de experimentele leergang terug.

## 3.2. Respons geven en krijgen op stelopdrachten

In de vorige paragraaf zijn enkele algemene processen bij collega-beoordeling aan de orde geweest. In deze paragraaf komen studies aan bod, waarin het becommentariëren en het commentaar verwerken door leerlingen op elkaars stelprodukten object van onderzoek was. Eerst stellen we de commentaarvaardigheid aan de orde, daarna bespreken we enkele onderzoeken waarin het accepteren en verwerken van commentaar centraal staat.

Er lijkt een zeker patroon in de ontwikkeling van commentaarvaardigheid vastgesteld te kunnen worden. Globaal verloopt dit van een natuurlijke lezersreactie via oppervlaktekenmerken van de tekst naar de inhoudelijke lijn en de inhoudelijke onderbouwing. De kwaliteit van de commentaren neemt toe naarmate leerlingen de taak vaker vervullen. In de commentaren zijn onderwijseffecten aan te wijzen, maar zij blijken ook natuurlijke lezersreacties te bevatten. Hoe nabijer het onderwerp van de tekst is, hoe meer leerlingen geneigd zijn andermans teksten 'invullend' te lezen en voorbij te gaan aan feilen.

Er zijn een paar onderzoeken waarin getracht is de ontwikkeling van commentaarvaardigheid vast te stellen. Hilgers (1984) vergeleek de commentaren van leerlingen van verschillende leeftijden op drie hun voorgelegde opstellen en op drie zelfgeschreven opstellen. Hij stelde vast dat oordelen over de eigen opstellen gedeeltelijk beïnvloed waren door oordelen die medeleerlingen al over dat werk hadden geveld. Hij schetst een ontwikkelingsverloop in vijf fasen, die kortweg als volgt getypeerd kunnen worden: identificeren met inhouden, beoordelen van oppervlaktekenmerken, beoordelen van de inhoud, beoordelen van de vaardigheid van de schrijver, beoordelen van de effectiviteit van de tekst. Een dergelijk natuurlijk ver-

loop dient volgens Hilgers (1984) via onderwijs bespoedigd te worden, omdat beoordelen het centrum van het schrijfproces zou zijn. Zo'n patroon lijkt zich ook voor te doen tijdens een stelcursus waarin leerlingenrespons wordt gebruikt (Ziv, 1983; Rubin, 1983). Via voornamelijk positieve reacties om een basis te scheppen voor groepsvorming komen inhoud en vorm steeds frequenter voor in commentaren. Al doende worden commentaren steeds duidelijker. Naarmate een cursus vordert worden meer suggesties voor verbeteringen gedaan (Ziv, 1983). Rubin (1983) demonstreert dat de reacties gaandeweg een andere opvatting over teksten tonen: eerst wordt gereageerd alsof een tekst een verzameling woorden is, later meer als een verzameling gedachten. Zij ondersteunt Hilgers' (1984) bewering dat het becommentariëren een onderwijseffect is, door te laten zien dat de commentaren steeds meer criteria uit de stelcursus gaan bevatten. Opmerkelijk is, dat een terugval naar een vorig niveau te zien is als een nieuwe tekstsoort in de cursus wordt geïntroduceerd. Rubin toont in een ander onderzoek (Rubin, 1983) weer de ontwikkeling van commentaarinhouden aan: van oppervlaktekenmerken, via globale inhoudelijke lijn naar gedetailleerde inhoudelijke opmerkingen. Danis (1982) onderzocht eveneens commentaren en concludeert dat het merendeel van de opmerkingen oppervlaktekenmerken betrof. Daarna volgde wat Danis noemde de categorie inhoud, waarin veel opmerkingen vielen over verheldering van onduidelijke woorden en zinnen en over de uitwerking van ideeën. Weinig opmerkingen werden gemaakt over de opbouw en de inperking van het onderwerp.

Newkirk (1984a, 1984b) verrichtte heel ander onderzoek. Hij vroeg leerlingen en docenten opstellen te beoordelen en vergeleek hun argumentaties. In beide onderzoeken blijkt dat leerlingen zich aangetrokken voelen tot opstellen waarin een thema werd behandeld waarmee zij zich sterk identificeerden. Dat had tot gevolg dat zij 'invullend' lazen, waardoor zij de door docenten bekritiseerde zwakke structuur niet opmerkten. Docenten willen juist dat de tekst het werk doet. Anderzijds bleken leerlingen minder flexibel de geleerde modellen toe te passen: men had geleerd een probleem van twee kanten te benaderen; een opstel waarin de schrijver zich vastbeet in één standpunt werd door leerlingen daarom qua argumentatie afgewezen, terwijl docenten dat juist wel waardeerden. Leerlingen leken gecharmeerd van een meer academisch register en het gebruik van stijlfiguren, terwijl docenten negatief oordeelden over de wijze waarop de stijlfiguren gebruikt waren. Wat we hier zien is dat leerlingen in twee rollen verkeren: lezer en leerlingcommentator. Zodra een tekst inhoudelijk boeit, overheerst de rol van de lezer. Is dat niet het geval, dan hanteert de leerling de schoolse, geleerde normen voor een goede tekst. Een docent vervult alleen de tweede rol. Hij neemt gedifferentieerder waar door de grotere kennis (bijv. de stijlfiguren). Zijn kennis stelt hem ook in staat een van het patroon afwijkende tekst te waarderen als een experiment. De verschillen tussen docenten en leerlingen zijn zelfs zo groot, dat docenten niet correct kunnen voor-

spellen wat het oordeel van leerlingen zal zijn (Newkirk, 1984b). Ziv (1983) ondervond dit aan den lijve. Opstellen die zij beoordeelde nadat deze herschreven waren, bekritiseerde zij soms op punten waarop zij juist herschreven waren op basis van de opmerkingen van medeleerlingen. Bovendien, de opstellen waren niet voor haar geschreven, zodat zij 'gaten' in de teksten aantrof, die voor de bedoelde lezers helemaal geen gaten waren, doordat zij veel meer een gemeenschappelijk referentiekader deelden met de schrijver.

Deze onderzoeken laten ons het probleem van de validiteit van het oordeel zien. Het inschakelen van leerlingen als commentatoren was een poging om het gebrek aan validiteit van docentoordelen op te vangen als het om communicatieve teksten voor medeleerlingen ging. Die medeleerlingen hanteren echter, als zij niet verleid worden tot thematische identificatie, zelf ook schoolse normen. Worden zij wel geboeid door de inhoud van de tekst, dan geven ze als hulpinstructeur die schrijver geen adequate feedback over meer schoolse kenmerken van het opstel. In paragraaf 3.4., Conclusies, komen we hierop terug.

Niet alle commentaar dat leerlingen krijgen wordt verwerkt in een herschreven tekst. Jones (1977) concludeert dat het afwijzen van commentaar in de helft van de gevallen terecht was. Er is dus ook commentaar dat, indien geaccepteerd en verwerkt, tot verbeteringen zou leiden. Welke processen spelen bij dit afwegingsproces een rol? Wellicht is een gemis aan revisievaardigheid een factor in dit proces. Rubin (1983) ontdekte dat leerlingen door leerlingenrespons wel een kritische leesvaardigheid ontwikkelden, die zij ook konden toepassen op eigen teksten, maar dat zij nog niet in staat waren de teksten op die punten te verbeteren. Ook andere factoren kunnen het accepteren en verwerken van commentaar beïnvloeden. Zo laat Danis (1982) zien dat het mondeling becommentariëren schrijvers voor problemen stelt. Schrijvers zijn onzeker over hun rol tijdens dit soort gesprekken, vertonen de neiging zich terughoudend op te stellen om de discussie snel af te ronden, weten niet wat ze van de anderen willen en vragen te weinig om verduidelijkingen.

Het kan ook zijn dat leerlingen de autoriteit van de feedbackgever in hun oordeelsproces betrekken. Soms wordt de instructiefout gemaakt leerlingenrespons toe te passen bij expressieve teksten. Danis (1982) constateerde dat schrijvers dan hun medeleerlingen niet als een valide publiek beschouwen, omdat er helemaal niet communicatief geschreven behoefde te worden. In dit verband vond Stone (1981), dat kritiek van medeleerlingen over inhoud en opbouw niet werd geaccepteerd: de schrijvers achtten zichzelf inhoudelijk expert. Kwam deze kritiek van de docent, dan werd deze wel verwerkt, terwijl zijn stilistisch commentaar terzijde werd gelegd: studenten denken dat de docent geen lezer maar leraar is, die fouten zoekt. Dergelijke fouten, opgemerkt door medeleerlingen werden wel verwerkt, blijkbaar omdat schrijvers denken dat deze oppervlaktekenmerken het begrip van de lezer beïnvloeden.

Ook Ziv (1983) concludeert dat schrijvers, ondanks hun vermogen tot commentaar geven, weinig verbeteringen aanbrengen in hun teksten als het om inhoudscommentaar ging. Zij veronderstelt dat leerlingen vinden dat 'peers' de inhoudelijke autoriteit missen en/of ervaring in het reviseren ontberen. Commentaar op het niveau van spelling, interpunctie, woordkeus etc. leidde wel tot revisies. Dit komt overeen met observaties van Jones (1977), die van 71 leerlingen van een highschool (seniors) de kladversie, de commentaarformulieren en de eindversie van drie schrijfopdrachten onderzocht. Leerlingen accepteerden en verwerkten vaker commentaar dan dat zij de kritiek naast zich neerlegden. Van de gevallen waarin commentaar werd afgewezen was dat in 50% terecht. In ruim 70% van de gevallen waarin kritiek werd verwerkt, leidde dat tot een betere tekst. Als schrijvers op eigen gezag verbeteringen aanbrachten, bleek dat in 15% van de gevallen een verslechtering, in 60% een verbetering. Gemiddeld bleven er per opstel nog 5 fouten ter correctie van de docent over. Meisjes verbeterden gemiddeld meer fouten in hun teksten op basis van commentaar dan jongens. De commentaren die tot verbeteringen leidden, konden ingedeeld worden in de volgende categorieën: woordkeus (30%), conventies (22%), zinsbouw (19%), inhoud (14%), alineabouw (10%) en structuur (6%). De uit eigen beweging aangebrachte verbeteringen vertoonden een ander frequentiepatroon: woordkeus (33%), inhoud (25%), zinsbouw (21%), conventies (11%), alineabouw (6%), structuur (4%). Kritiek die afgewezen werd, had veelal betrekking op perspectief, grammaticale zaken, alineabouw en inhoud.

Een laatste factor kan een eigenschap van de persoon van de feedbackontvanger zijn. Slechte schrijvers veranderen andere zaken in hun tekst dan goede schrijvers, omdat zij andere opvattingen over een goede tekst hebben (Stone, 1981). De persoonlijkheid en het eigen oordeel over becommentarieerde tekst speelt evenzeer een rol in de wijze waarop commentaar verwerkt wordt (Berkenkotter, 1984). Op beide onderzoeken gaan we kort in. Stone (1981) onderzocht de herschrijfgewoonten van eerstejaars- en gevorderde studenten, na commentaar van medestudenten. In zijn onderzoek betrok hij 150 leerlingen, waarover hij impressionistisch materiaal verzamelde. Hij concludeert dat herschrijven meer zin heeft, naarmate leerlingen gevorderder zijn, omdat alleen zij zich bezig houden met inhoud en opbouw en zich niet tot oppervlaktekenmerken beperken. Leerlingen met lage cijfers bleken vooral hun slechtste opstellen te willen verbeteren, terwijl leerlingen met hoge cijfers liever de betere opstellen herschreven. Stone veronderstelt dat zwakkere leerlingen denken dat herschrijven het verbeteren van oppervlaktestructuren is en dat slechte opstellen nu eenmaal veel van dergelijke fouten bevatten. Dit verschijnsel deed zich ook voor toen gevorderder studenten werd gevraagd één van hun zes opstellen te herschrijven. Stone onderzocht tevens waarin de herschrijvingen van goede en minder goede leerlingen verschilden. Vrijwel iedereen voegde

inhoud toe: bij goede leerlingen waren dit werkelijk verbeteringen, slechte leerlingen vervingen voorbeelden door andere voorbeelden, wat geen verbetering was. Goede leerlingen veranderden met succes de opbouw. Slechte leerlingen veranderden meer woorden dan goede leerlingen, maar de veranderingen van deze laatsten waren vaker verbeteringen. Slechte leerlingen plakten wat zinnen via nevenschikking aan elkaar, terwijl betere leerlingen vaker transformaties toepasten als onderschikking, inversie, parallellie. Vooral qua zinsbouw verbeterden de opstellen van goede leerlingen het meest.

Berkenkotter (1984) onderzocht bij 10 eerstejaarsstudenten hoe zij zich gedroegen in commentaargroepen en wat er van het commentaar tot revisies werd verwerkt. Zij analyseerde daartoe hardopschrijfprotocollen van diverse kladversies en geluidsbanden van groepsinteracties. Ze concludeert dat de wijze waarop schrijvers omgaan met commentaar sterk afhangt van de persoonlijkheid van de schrijver, zijn ontwikkelingsniveau en zijn vaardigheid om schrijfproblemen aan te pakken. Aan de hand van drie gevallen wordt deze conclusie toegelicht. Deze drie gevallen verschillen sterk in de mate waarin zij zeggenschap hebben over hun tekst. De eerste leerling is buitengewoon kritisch over andermans teksten, maar laat kritiek op de eigen teksten niet toe, met als argument dat het zíjn mening is, zíjn tekst. Een tweede schrijver krijgt zeer positief commentaar, maar is toch niet tevreden over zijn tekst en verbetert deze grondig, onafhankelijk van het lezerscommentaar. Het derde voorbeeld doet haar uiterste best om tegemoet te komen aan het commentaar, maar bemerkt pas veel later dat die opmerkingen weinig grond hadden.

## 3.3. Effectstudies

In deze paragraaf bespreken we een groep van 21 effectstudies, op een viertal aspecten. We doen dat met het oog op een mogelijke bijdrage aan de theorievorming, constructie van het experimentele onderwijsprogramma, het onderzoeksontwerp en interessante afhankelijke variabelen. Het zijn vrijwel alle zogenaamde quasi-experimentele onderzoeken. Onderzoeken met een meer impressionistisch karakter zoals Smith (1975) en Steinacker (1976), zijn in hoofdstuk 2. verwerkt. In de meeste gevallen heeft men zowel een voor- als een natoets afgenomen, zodat de onderzoeksgroepen vergeleken kunnen worden bij de start van het onderzoek. Omdat echter de proefpersonen niet via loting aan de condities zijn toegewezen, spreekt men van quasi-experimenten (Cook en Campbell, 1979, p.6). Bij analyse-problemen in dit soort studies zullen we kort stilstaan. We lichten de onderzoeken achtereenvolgens op vier thema's door. Welke onafhankelijke variabelen zijn in verband gebracht met leerlingenrespons? (tabel 3.1.). Op welke afhankelijke variabelen van schrijfvaardigheid verwachtte men effecten van leerlingenrespons aan te tonen (tabel 3.2.)? Wat zijn de

statistische merites van de onderzoeken (tabel 3.3.)? Tenslotte: zijn er aanwijzingen voor het ontwerp van het experimentele onderwijsprogramma te ontlenen aan de succesvol gebleken experimentele onafhankelijke variabelen?

Heeft het inschakelen van leerlingen als commentator van opstellen meer effect op schrijfvaardigheid dan wanneer de docent commentaar geeft? Uit tabel 3.1. blijkt dat men slechts in acht van de 21 onderzoeken erin geslaagd is statistisch significante verschillen aan te tonen tussen scores van de verschillende groepen. Van die onderzoeken tonen zeven een positief resultaat ten gunste van de experimentele didactiek, één, Earls (1983) vond de controledidactiek superieur. Voor attituden jegens schrijven of schrijfangst is het beeld niet gunstiger: in drie van de tien onderzoeken kon een significant verschil ten gunste van de experimentele didactiek vastgesteld worden.

Tabel 3.1.: De variabelen die in 21 effectstudies gerelateerd zijn aan schrijfvaardigheid en attituden t.a.v. schrijven of schrijfangst, met tussen haakjes het aantal malen dat de relatie onderzocht is. De getallen verwijzen naar de nummers van de studies die in bijlage 2 vermeld zijn.

| Variabelen | | Schrijfvaardigheid | | | | Attituden/Schrijfangst | |
|---|---|---|---|---|---|---|---|
| | | niet getoetst | geen significant effect | significant effect | | geen significant effect | significant effect |
| Onderwijskenmerken | | | | | | | |
| 1. Onderwijsprogramma | (21) | 2 | 3,4,6,7,9, 11,14,16,17,18 20,21 | 1,5,8,10,12 13,15,19 | (10) | 1,3,7d,9,14, 15,17,20 | 7d,11,21 |
| 2. Docent | (8) | | 4,5,6,11,14,18 | 1,11a,20 | (4) | 11,14c,20e | 1,14c,11a 20e |
| 3. Klas | (2) | | 6,17 | | | 17 | |
| Leerlingkenmerken | | | | | | | |
| 4. Sexe | (3) | | 4,19b | 1,19b | (2) | 20 | 1 |
| 5. Intro-extraversie | (2) | 14,17 | | | | | |
| 6. Schrijfangst | | | 11 | | | | |
| Interacties | | | | | | | |
| 7. prog. x doc. | (5) | | 5,11,18,20 | 1 | (3) | 1,11 | 20 |
| 9. prog. x sexe ll. | (4) | | 11 | 1,9,16,19 | (2) | | 1,9 |
| 9. doc. x sexe ll. | (2) | | 1,4 | | (2) | 1 | 20 |

a: Fox (1979) : docenteffect treedt niet op bij analyses op de gehele groep; wel bij analyses op subgroep (hoog-schrijfangstigen).
b: Sager (1973) : sexe-effect op drie subvariabelen, op twee andere niet.
c: Lyons (1976) : docenteffect treedt op voor 1 subvariabele van attituden; bij drie andere subvariabelen niet.
d: Deluney (1980) : voor sommige attitude variabelen wel, voor andere geen significant effect.
e: Sears (1971) : docenteffect treedt op voor inzetverwachtingen, niet voor schatting van eigen bekwaamheid.

Van de variabelen in de categorie Onderwijskenmerken, is vooral de variabele docent belangrijk. Een aantal onderzoeken heeft men zo opgezet dat één docent aan twee klassen lesgeeft: een experimentele en een controleklas. In de onderzoeken van Benson (1979), Fox (1979), Lyons (1976) en Sears (1971) blijkt de variabele docent een statistisch significant effect teweeg gebracht te hebben. In het onderzoek van Benson (1979) is het effect spurieus, omdat de variabele docent samenvalt met leerjaar en leeftijd van leerlingen met verschillen tussen scholen. In het onderzoek van Fox (1979) blijkt het effect van de variabele docent alleen als de gegevens van een subgroep, namelijk de hoogschrijfangstige leerlingen, geanalyseerd worden. Het is niet onaannemelijk dat vooral deze groep leerlingen gevoelig is voor de persoon van de docent. Bij Lyons (1976) werd het docenteffect slechts aangetoond voor één attitude concept, 'showing my writing to teachers', begrijpelijk een docentafhankelijk concept. In het onderzoek dat Sears (1971) verrichtte, toonde zij zowel voor schrijfprestaties als voor een attitude concept, - de schatting van de inzet -, een docenteffect aan. Voor een ander attitude concept, - de schatting van de eigen bekwaamheid -, werd geen docenteffect gevonden. Kortom, als het om attituden gaat, is een docenteffect in drie onderzoeken aangetoond, terwijl in dezelfde onderzoeken voor andere concepten van attituden geen docenteffect geconstateerd werd. Sommige attituden lijken dus docentgevoelig. Als het om schrijfvaardigheid gaat lijkt een docenteffect niet gauw te verwachten, tenzij het een gevoelige groep leerlingen betreft (Fox, 1978).

De leerlingkenmerken die onderzocht en getoetst zijn, zijn sexe, mate van intro- en extraversie en schrijfangst. Benson (1979) vond dat meisjes beter presteerden dan jongens; Carter (1982) vond eenzelfde tendens, maar niet statistisch significant. Sager (1973) vond voor een drietal, - hoog correlerende -, variabelen een significant effect voor sexe (ondersteunende informatie, zinsstructuur en globaal totaaloordeel), voor twee andere variabelen niet (woordkeuze, opbouw). Benson (1979) trof een effect van sexe aan voor zowel opstelkwaliteit, attitude, lengte van opstellen en revisie op alineaniveau, alle ten gunste van meisjes. Als er dus een effect van sexe gevonden wordt, blijkt het steeds te gaan om meisjes die beter presteren dan jongens.

Fox (1978) toonde aan dat schrijfangstige leerlingen significant beter presteerden na Fox' experimentele onderwijsprocedure dan de schrijfangstige leerlingen die onderwijs volgden in de controleconditie. Hierboven bleek al dat het erop lijkt dat deze groep leerlingen gevoeliger is voor de persoon van de docent; dit lijkt ook voor het gevolgde onderwijs te gelden.

Van de onderzochte interacties is die tussen programma en andere variabelen interessant. Nu blijkt dat de interactie tussen docent en programma van de acht keer maar twee keer

significant is, mogen we verwachten dat docenten geen systematisch betere prestaties veroorzaken op een van beide programma's die zij onderwijzen. Opvallend is het relatief groot aantal keren dat een significante interactie tussen programma en sexe van leerlingen werd aangetoond, namelijk vijf van de zes keer. Dit zou erop kunnen duiden dat de onderwijsprogramma's sexespecifiek zijn. In het onderzoek van Benson (1979) bleken meisjes baat te hebben bij gestructureerde, informatieve leerlingenfeedback, terwijl jongens beter presteerden in de conditie van docentfeedback. Bij Myers (1979) bleken jongens en meisjes in de controleconditie niet te verschillen, maar presteren meisjes in de experimentele conditie significant beter. Farrell (1977) meldt een daaraan tegengesteld resultaat. Sager (1978) rapporteert een nivellerend effect van het experimentele onderwijsprogramma: de verschillen tussen jongens en meisjes werden in de experimentele conditie kleiner. De jongens hadden baat bij leerlingenrespons. Hoewel we erop verdacht moeten zijn dat een dergelijke interactie zich voor kan doen, is er te weinig houvast om specifieke verwachtingen te koesteren over een dergelijke interactie.

Een conclusie van heel andere aard is deze. Uit tabel 3.1. kan opgemaakt worden dat de verslaglegging van de geraadpleegde onderzoeken gebrekkig is. Soms worden resultaten vermeld van het hoofdeffect docent, maar is niets te vinden over interactie tussen docent en programma (4,6,14, zie tabel 3.1.). In een onderzoek waarin sexe is opgenomen als variabele, wordt niets vermeld over een interactie tussen sexe en programma (4 bijvoorbeeld). Het omgekeerde komt ook voor: wel wordt gewag gemaakt van een interactie tussen sexe en programma, terwijl niets vermeld wordt over een hoofdeffect van sexe (11). Helaas heeft maar één onderzoeker (Fox, 1979) gebruik gemaakt van de mogelijkheid gegevens te analyseren voor subgroepen. Anders was er meer kennis geweest over de invloed van het niveau van schrijfvaardigheid: is leerlingenfeedback vooral voor goede of vooral voor slechte leerlingen een goede onderwijsprocedure?

Een tweede analyse van de onderzoeken betrof een beschouwing van de afhankelijke variabelen binnen schrijfvaardigheid. In een aantal onderzoeken heeft men namelijk aspecten van stelvaardigheid onderscheiden. Wellicht was het mogelijk veelbelovende subvariabelen op het spoor te komen. In tabel 3.2. staat een overzicht.

De gegevens in tabel 3.2. laten zien dat men vooral verwachtte op formele kwaliteiten verschillen te vinden. In tien onderzoeken zocht men naar kwaliteitsverschillen binnen de categorieën Spelling en Interpunctie en Formulering en Stijl, in drie ervan met succes. Naar verschillen in Opbouw en Inhoud is veel minder vaak gezocht. Op geen van de in tabel 3.2. genoemde variabelen is opvallend vaak een significant verschil gevonden.

Tabel 3.2.: Specifieke Schrijfprestatievariabelen in effect-studies. De getallen verwijzen naar de studies in bijlage 2

| Categorie | Subcategorie | |
|---|---|---|
| Spelling en interpunctie | - spelfouten | 7 |
| | - interpunctie fouten | 3,7,16 |
| | - kennis van spelling en interpunctie | 5*,8,15* |
| | - kennis van grammatica | 5,10,15* |
| Formulering en stijl | - woordkeus | 19* |
| | - formuleerfouten | 7 |
| | - correcte, gevarieerde zinsbouw | 7,15* |
| | - goede formulering | 19* |
| | - zinsbeginnen | 2 |
| | - aantal woorden per T-unit | 1,3,20 |
| Opbouw | - duidelijke lijn | 7,15* |
| | - logische opbouw | 7,16,19* |
| Inhoud | - ondersteuning, uitwerking | 7,19* |
| | - duidelijk thema | 7 |
| | - belangwekkend | 7 |
| | - consistent perspectief | 7 |
| Lengte | - | 1*, 11 |
| Revisie | - totaal | 1* |
| | - kosmetisch | 1* |
| | - woord | 1* |
| | - zin | 1* |
| | - alinea | 1* |

* = significant ($p \leq .05$)

Er komt geen veelbelovende afhankelijke variabele naar voren. De nadruk die de onderzoekers leggen op de formele aspecten, duidt erop dat men verwacht dat door leerlingenfeedback vooral de editorfunctie verbetert. De processtudies uit par. 3.2. laten zien dat er wel veel commentaar op deze variabelen overgenomen wordt door de schrijver, maar blijkbaar is dit effect ook werkzaam in de groepen die docentfeedback krijgen of transfereert deze kennis niet naar nieuwe taken. In één onderzoek (Benson, 1979) werd onderzocht of er meer gereviseerd werd als gevolg van leerlingenrespons. Het succes van dit onderzoek past in de verwachtingen met betrekking tot revisie: goede schrijvers reviseren meer en op een hoger niveau dan slechte schrijvers.

Uit tabel 3.3. blijkt dat vrijwel alle onderzoeken volgens een voortoets-natoets ontwerp zijn opgezet. Het voordeel van het opnemen van een voortoets in het ontwerp is de mogelijkheid verschillen in startbekwaamheid bloot te leggen. Enkele onderzoeken zijn echte experimenten, d.w.z. dat de proefpersonen via loting aan condities zijn toegewezen met het oog op bestrijden van selectie-effecten (Burt, Carter, Clifford, Fox,

Maize, Myers, Pierson). De andere onderzoeken zijn quasi-experimenteel van opzet: intacte groepen (klassen) werden in hun geheel aan een conditie toegewezen. Meestal gebeurde dat volgens het lot. Zou men in dat geval klassen als observatie-eenheden beschouwen, dan is ook sprake van een experimentele opzet. Men analyseert echter steeds op leerlingniveau. Er is dan sprake van een quasi-experimentele opzet: men heeft dan wel enige controle op selectie-effecten uitgeoefend, maar het selectie-effect van de klassenindeling is niet gecontroleerd. Deze quasi- of semi-quasi-experimentele opzet heeft het nadeel dat de kans groter is dan in een experimentele opzet dat de condities al bij aanvang van het experiment van elkaar verschillen op relevante variabelen, zoals de mate waarin de afhankelijke variabele aanwezig is. In de fase van statistische analyse kan voor deze ongelijke startpositie gecorrigeerd worden met behulp van covariantieanalyse.

Naast het probleem van selectie-effecten, is er in een dergelijke onderzoeksopzet het probleem van het analyseniveau omdat men de leerling als onderzoekseenheid definieert. In echte experimenten deelt men proefpersonen via loting in condities in: het analyseniveau is dan terecht het individuele niveau. Het voordeel hiervan is dat het aantal ongecorreleerde waarnemingen groot is en daarmee het onderscheidingsvermogen van het experiment. In quasi-experimenten wijst men echter zoals gezegd klassen aan condities toe. Het adequate analyseniveau zou in dat geval de klas zijn, wat een aanzienlijke afname aan vrijheidsgraden en van het onderscheidingsvermogen met zich brengt. In geen van de geanalyseerde onderzoeken is dit probleem onderkend en analyseerde men op individueel niveau alsof er in klassen niet een klasse-effect bestaat, een correlatie tussen de individuele scores. Als men uitgaat van een middelmatige effectgrootte ($f=0.25$) en een intraclasscorrelatie van .07, zou men ongeveer tweederde van het aantal vrijheidsgraden verliezen. De consequentie voor het onderscheidingsvermogen van de onderzoeken in tabel 3.3. is duidelijk. Dit probleem klemt des te meer in het geval van een didactiek zoals in deze studies onderzocht wordt. Deze didactiek wordt namelijk niet op individueel niveau uitgeoefend, maar veeleer op klasseniveau, zeker in de experimentele didactiek van leerlingenrespons met alle mogelijkheden om met, van en door elkaar te leren.

De laatste kolommen van tabel 3.3. bevatten schattingen van het onderscheidingsvermogen van de experimenten, gebaseerd op Cohen (1977): de waarschijnlijkheid waarmee de nulhypothese verworpen zal kunnen worden. De schatting is gemaakt voor $\alpha=.05$ en een middelmatige effectgrootte ($f=0.25$), dat wil zeggen de mate waarin het fenomeen waarop het onderzoek zich richt, zich in de werkelijkheid voordoet. Deze effectgrootte komt overeen met een intraclasscorrelatie van .24, dat wil zeggen dat bijna 6% van de variantie binnen een populatie verklaard wordt door het lidmaatschap van deelpopulaties (Cohen, 1977, p.26 e.v.). Uit de laatste kolom blijkt dat maar weinig onderzoeken een kans groter dan vier op vijf hadden om de nulhypothese te verwerpen. Alleen Benson (1979), Myers (1979) en Ward (1959) halen deze norm. Dit betekent dat op de

Tabel 3.3.: Statistische gegevens van 21 (quasi-)experimentele onderzoeken. In de laatste kolommen is het onderscheidingsvermogen vermeld voor een middelmatig effect ( $\alpha$ =.05)

| Onderzoek | Afhankelijke variabele | Voortoets | Natoets | Betrouwbaarheid |
|---|---|---|---|---|
| Benson | 1. attitude | 13 items uit de Writing Apprehension Test | idem | - |
| | 2. schrijfvaardigheid | 20 min.opstel | idem | voortest .63<br>natest .73 |
| | 3. revisievaardigheid | - | opstel van posttest | |
| Bouton | 1. Variaties in zinsbeginnen | ? | ? | |
| Burt | 1. schrijfvaardigheid | brief, 45 min., T-unittelling | idem | - |
| | 2. interpunctie | interpunctiefouten in brief | idem | |
| Carter | 1. schrijfvaardigheid | 90 min.opstel | idem | holistisch .69-.94<br>analytisch .81-.94 |
| Clifford | 1. schrijfvaardigheid | opstel | idem | .79 (jury) |
| | 2. grammaticale kennis | schrijftest | idem | .84 |
| | 3. grammaticale vaardigheid | foutscore | idem | .83 |
| Copland | 1. schrijfvaardigheid | 2 opstellen, 45 min. elk | idem | .81 |
| Delaney | 1. schrijfvaardigheid | opstel 300 à 500 woorden | idem | .84 |
| | 2. syntactische ontwikkeling | Hunts T-unit test | idem | - |
| | 3. attitude | Osgood's attitutest | idem | - |
| Earls | 1. schrijfvaardigheid | essay, 90 min. | idem | .76 |
| | 2. conventies | STEP test 2A | idem | - |
| Farrell | 1. schrijfvaardigheid | opstel, 45 min. | idem | voortest .70<br>natest .80 |
| | 2. conventies | STEP test 2A | idem | .91 |
| | 3. attituden | Farrells att. vragenlijst | idem | .99 |
| Ford | 1. grammaticale kenm. | schrijftest | idem | - |
| | 2. schrijfvaardigheid | opstel | idem | - |

vervolg tabel 3.3.

| Analysetechniek | n ppn | n doc | n cond. | onderscheidingsvermogen doc. | onderscheidingsvermogen condities |
|---|---|---|---|---|---|
| Ancova | 288 | 6 | 3 | 92 | 97 |
| Ancova | 288 | 6 | 3 | 92 | 97 |
| Anova | 288 | 6 | 3 | 92 | 97 |
| Ancova[1)] | 56 | ? | 2 | - | 46 |
| Ancova[1)] | 87 | 1 | 2 | - | 64 |
| Ancova[1)] | 87 | 1 | 2 | - | 64 |
| Ancova[1)] | 65 | 2 | 2 | 53 | 53 |
| Ancova | 92 | 3 | 2 | 56 | 67 |
| Ancova | 92 | 3 | 2 | 56 | 67 |
| Ancova | 92 | 3 | 2 | 56 | 67 |
| Ancova | 98 | 2 | 2 | 70 | 70 |
| Ancova | 46 | 1 | 2 | - | 39 |
| Ancova[1)] | 46 | 1 | 2 | - | 39 |
| Ancova[1)] | 46 | 1 | 2 | - | 39 |
| Ancova | 81 | 2 | 2 | 61 | 61 |
| Ancova | 81 | 2 | 2 | 61 | 61 |
| Ancova | 154 | 3 | 4 | 80 | 73 |
| Ancova | 154 | 3 | 4 | 80 | 73 |
| Ancova | 154 | 3 | 4 | 80 | 73 |
| Ancova | 50 | 1 | 2 | - | 42 |
| Ancova | 50 | 1 | 2 | - | 42 |

vervolg tabel 3.3

| Onderzoek | Afhankelijke variabele | Voortoets | Natoets | Betrouwbaarheid |
|---|---|---|---|---|
| Fox | 1. attituden | W.A.T. | idem | |
| | 2. schrijfvaardigheid | - | opstel, 90 min. | .92(.64) |
| | 3. lengte | - | aantal woorden woorden | - |
| | 4. schrijfvaardigheid bij hoogschrijfangstigen | | opstel, 90 min. | |
| | 5. lengte van opstellen bij hoogschrijfangstigen | | aantal woorden | |
| Karegianes | 1. schrijfvaardigheid | opstel | idem | voortest .87<br>natest .85 |
| Lagana | 1. kennis over schrijven | Step Writing Test | idem | - |
| | 2. schrijfvaardigheid | Step Essay Test | idem | - |
| Lyons | 1. schrijfvaardigheid | 2 opstellen (2x10 min.) | idem | .98 |
| | 2. attituden | Lyons' Semantische differentiaal (4x) | idem | .90-.96 |
| Maize | 1. schrijfvaardigheid | opstel | idem | - |
| | 2. interpunctie, hoofdletters | schrijftest | idem | |
| | 3. dictie, verwijzingen | schrijftest | idem<br>idem | |
| | 4. zinsbouw, logische opbouw | schrijftest | idem | |
| Myers | 1. alineabouw | opstel (200 woorden) | idem | - |
| | 2. interpunctie | idem | idem | - |
| Pfeifer | 1. schrijfangst | - W.A.T.<br>- Pfeifers Writing Apprehension Measure | idem<br>idem | .93<br>.87 |
| | 2. schrijfvaardigheid | opstel (45 min.) | idem | |
| Pierson | 1. schrijfvaardigheid | Step Writing test | idem | |

vervolg tabel 3.3

| Analysetechniek | n ppn | n doc | n cond. | onderscheidingsvermogen doc. | condities |
|---|---|---|---|---|---|
| Ancova | 107 | 3 | 4 | 62 | 54 |
| Anova | 107 | 3 | 4 | 62 | 54 |
| Anova | 107 | 3 | 4 | 62 | 54 |
| t.test | 60 | 3 | 2 | 33 | 47 |
| t.test | 60 | 3 | 2 | 33 | 47 |
| Ancova | 49 | 1 | 2 | - | 40 |
| Ancova[1)] | 60 | 1 | 2 | - | 49 |
| Ancova[1)] | 60 | 1 | 2 | - | 49 |
| Ancova | 106 | 2 | 2 | 73 | 73 |
| Ancova | 106 | 2 | 2 | 73 | 73 |
| Ancova[1)] | 149 | ? | 2 | | 55 |
| Ancova[1)] | 149 | ? | 2 | | 55 |
| Ancova[1)] | 149 | ? | 2 | | 55 |
| Ancova[1)] | 149 | ? | 2 | | 55 |
| Ancova | 168 | 4 | 2 | 78 | 90 |
| Ancova | 168 | 4 | 2 | 78 | 90 |
| Ancova | 92 | ? | 2 | | 67 |
| Ancova | 92 | ? | 2 | | 67 |
| Ancova | 92 | ? | 2 | | 67 |
| Ancova[1)] | 153 | 3 | 2 | 80 | 87 |

vervolg tabel 3.3

| Onderzoek | Afhankelijke variabele | Voortoets | Natoets | Betrouwbaarheid |
|---|---|---|---|---|
| Sager | 1. schrijvaardigheid | opstel | idem | .97 |
| Sears | 1. schrijfvaardiheid | (leesscores als covariant) | opstel | .63 |
| | 2. attituden: schatting eigen inzet | eigen instrument (24 items) | idem | .76[2)] |
| | 3. attituden: schatting van eigen kunnen | eigen instrument (7 items) | idem | .88[2)] |
| Ward | 1. schrijfvaardigheid | opstel | idem | |
| | 2. kennis over schrijven | Step Writing test | idem | |
| | 3. attitude jegens cursus | Ward's schaal | idem | |

1) In een aantal onderzoeken is geen covariantie-analyse gebruikt als analysetechniek, maar variantie-analyse (Burt, Carter), t-testen (Delaney) en verschilscores (Lagana, Maize, Ward). De onderzoeksopzetten laten echter covariantie-analyse wel toe. Ter wille van de vergelijkbaarheid is het onderscheidingsvermogen toch op basis van covariantie-analyse berekend.

2) Test-hertestcorrelatie

resultaten van de overige studies weinig staat gemaakt kan worden als géén statistisch significant verschil gevonden werd tussen de experimentele conditie en de controleconditie. Als men, zoals Delaney, slechts 39% kans heeft om een statistisch significant resultaat aan te treffen en men vindt het niet, is dan het bewijs geleverd dat de experimentele didactiek geen winst boekte vergeleken met de controledidactiek? Wat men in een dergelijk geval doet, is: uitermate streng zijn in het kiezen van een significantieniveau van .05, - 95% zekerheid dat de verschillen niet aan kans zijn toe te schrijven -, en zeer coulant zijn tegenover het lage onderscheidingsvermogen. Waarom zou men onderzoeken of een didactiek vruchten afwerpt als men het onderzoek zo inricht dat men slechts een kans van 0,40 heeft de verschillen te kunnen aantonen als ze bestaan? Wat men zou kunnen doen, hoewel dat niet gebruikelijk is, is het significantieniveau wat minder streng te nemen ( $\alpha$ =.10), waardoor het onderscheidingsvermogen toeneemt (bij Delaney bijv. tot 0,53).

vervolg tabel 3.3.

| Analysetechniek | n ppn | n doc | n cond. | onderscheidingsvermogen doc. | onderscheidingsvermogen condities |
|---|---|---|---|---|---|
| Ancova | 63 | 3 | 2 | - | 62 |
| Ancova | 117 | 4 | 2 | 60 | 77 |
| Ancova | 117 | 4 | 2 | 60 | 77 |
| Ancova | 117 | 4 | 2 | 60 | 77 |
| Ancova[1)] | 107 | 1 | 2 | - | 73 |
| Ancova[1)] | 107 | 1 | 2 | - | 73 |
| Ancova[1)] | 107 | 1 | 2 | - | 73 |

De conclusie die we trekken uit het overzicht van de statistische eigenschappen van de 21 geanalyseerde onderzoeken, is dat de onderzoeksopzetten en statistische bewerkingen te wensen overlaten. Het is niet uitgesloten dat veel van de onderzoeken 'ten onrechte' geen verschillen tussen de onderwijsprogramma's constateerden, als gevolg van een inadequaat statistisch ontwerp.

Een vierde analyse van de effectstudies betrof de onafhankelijke variabele: het onderwijs zoals dat in de onderzoeken is vormgegeven. Een confrontatie van succesvolle en niet succesvolle onderwijsprogramma's zou mogelijk werkzame bestanddelen van succesvolle programma's blootleggen. Deze hoop bleek ijdel. De onderzoeken lieten zich gewoonweg niet op een aantal variabelen beschrijven. Ten eerste bleken de controle en experimentele onderwijsprogramma's vaak op meer kenmerken te verschillen dan de feedbackfase. Zo waren er verschillen in de

instructiefase (Copland, 1980; Clifford, 1981; Delaney, 1980; Fox, 1978; Lagana, 1972; Maize, 1954; Sager, 1973), in het aantal opstellen dat geschreven werd (Lagana, 1972; Maize, 1954), in de mate van individualisering van het programma (Delaney, 1980; Lagana, 1972), het al dan niet herschrijven van opstellen (Clifford, 1981; Earls, 1983; Fox, 1978). In sommige onderzoeken was er zelfs in de controleconditie in zekere mate sprake van het uitwisselen van opstellen of het voorlezen van opstellen aan elkaar (Delaney, 1980; Lyons, 1976; Sager, 1973). Wilden we een analyse verrichten op de succesvolle aanpakken, dan moesten we dus een aantal studies elimineren. We vergeleken de aanpakken van zeven studies, waarvan er drie succesvol waren (Benson, 1979; Karegianes et al., 1980; Ford, 1973) en vier in een onbeslist eindigden (Carter, 1982; Farrell, 1977; Myers, 1979; Pfeifer, 1981). In al deze studies werd gebruik gemaakt van leerlingenrespons in pure vorm. In alle gevallen werd gebruik gemaakt van commentaarformulieren. Alleen in Myers (1979) werd instructie gegeven over tekstkwaliteiten; in de andere studies bestond de sequentie van lesactiviteiten uit schrijven, becommentariëren en herschrijven. Alleen bij Farrell (1977) werd ook tijdens de ontwerpfase van opstellen overleg gevoerd tussen leerlingen. Geen verschillen tussen de twee groepen studies vonden we op de variabelen: wijze van feedbackgeven, keuze, aantal en anonimiteit van feedbackgevers, hoeveelheid instructie, moment van feedbackgeven en de verplichting tot herschrijven. Evenmin bleek een verschil tussen de leeftijden van de proefpersonen. We slaagden er kortom niet in werkzame bestanddelen te achterhalen.

De resultaten van de vier analyses brengen ons niet veel verder. Vooralsnog krijgt de theorie uit hoofdstuk 2 weinig steun van onderzoeksresultaten. Vooral ten aanzien van attituden en schrijfangst lijken de praktijkuitspraken overtrokken. Attituden en schrijfangst zijn wellicht vrij stabiele kenmerken en laten zich niet gemakkelijk door stelonderwijs veranderen. Het onbesliste karakter van de tweekamp tussen leerlingenrespons en docentfeedback kan toegeschreven worden aan een ongelukkige keuze van afhankelijke variabelen. Uit de keuze van specifieke schrijfprestatievariabelen in de onderzoeken blijkt dat men vooral effecten verwacht van de leerling als onderwijshulp en niet van de leerling als lezer; in dat geval had men afhankelijke variabelen op meer retorisch niveau gekozen. Het uitblijven van een beslissing in het voordeel van leerlingenrespons kan ook toegeschreven worden aan de gebrekkige statistische ontwerpen.

### 3.4. Conclusies

In dit hoofdstuk hebben we drie groepen van onderzoeken besproken om meer kennis op te doen voor het samenstellen van een stelcursus op basis van leerlingenrespons. In deze para-

graaf trekken we onze conclusies uit deze literatuur: de algemene literatuur over storende factoren in collega-beoordelingsprocessen, de literatuur over dergelijke processen in stelonderwijs en de effectstudies over leerlingenresponsexperimenten. Uit de effectstudies trekken we tevens conclusies over het ontwerp van het experiment.

Algemene literatuur. Het geven en ontvangen van collega-beoordelingen wordt beïnvloed door affectieve relaties, sexe en 'image-building'. Het lijkt wenselijk suggesties uit de algemene beoordelingsliteratuur over te nemen. Het is beter als de beoordelaars de schrijver niet kennen en als zij beoordelen op basis van criteria. Beoordelaarsgroepen zouden van een gemengde samenstelling moeten zijn om sexespecifieke feedbacktendenties te spreiden. Het is beter de beoordelingen die leerlingen verrichten niet door de docent te laten beoordelen, wat door sommige voorstanders van leerlingenrespons in onderwijs, zoals Bruffee (1980), wel bepleit wordt. De feedback moet onpersoonlijk en atomistisch zijn, omdat leerlingen dit bruikbare feedback vinden. Als het om taken gaat, die erg specifiek zijn, moet de beoordelaar competent zijn op dat terrein. Dat zou voor gedocumenteerde opstellen betekenen, dat de beoordelaar-leerling inhoudelijk op het terrein thuis is, doordat hij bijvoorbeeld zelf een opstel schreef over dat onderwerp.

Collega-beoordeling kan door het openbare karakter van de beoordelingssituatie leiden tot een toename in motivatie (House, 1980; Salili et al., 1976). Deze studies ondersteunen de theorie (hoofdstuk 2). In de effectstudies (par. 3.3.) blijkt dit echter niet op te gaan voor aan schrijven gebonden attituden. Men moet voorkomen dat leerlingen zich conformeren, - onder druk -, aan de oordelen van anderen (White, 1972; Silver en Schlenker, 1981) omdat dat belemmert dat leerlingen zelf een normen- en waardensysteem opbouwen, wat juist bij stellen zo belangrijk is, aangezien leerlingen tijdens en na het schrijven hun teksten moeten herzien met kritische blik. Het toevoegen van een opdracht het opstel eerst zelf te beoordelen alvorens het uit handen te geven, zou het gevaar van een te grote mate van afhankelijkheid van anderen kunnen bestrijden.

Als verscheidene personen feedback geven, is de kans groot dat de oordelen uiteenlopen. Leerlingen die geconfronteerd worden met uiteenlopende oordelen over hun tekst, zullen zelf te gunstig oordelen over die tekst. Het gevolg kan zijn dat zij minder grondig zullen herschrijven dan op basis van het gemiddelde van de beoordelingen gewenst zou zijn. Er is wat voor het standpunt te zeggen dat veel variantie in oordelen schrijvers juist te denken zou moeten geven: blijkbaar is de tekst niet eenduidig gepercipieerd. Juist dan zouden schrijvers heel kritisch de oordelen en oorzaken van oordelen moeten analyseren en niet moeten vluchten in een overschatting van hun eigen oordeel. In het experimentele onderwijsprogramma zou

dit voorkomen moeten worden. Twee oplossingen zijn denkbaar. Men zou alleen het gemiddelde van oordelen aan schrijvers kunnen terugkoppelen. Nadeel van deze oplossing is dat zo'n score de werkelijke perceptie van lezers niet weergeeft. De schrijvers hebben wel minder gelegenheid wat gemakkelijk over de eigen tekst te denken, maar het didactische element van diversiteit in lezers gaat verloren. Een tweede oplossing kan gerealiseerd worden door in de herschrijfinstructies op te nemen dat schrijvers met de twee laagste van bijvoorbeeld drie oordelen rekening moeten houden als ze gaan herschrijven. Het is de vraag of een dergelijke oplossing geen negatieve bij-effecten met zich meebrengt. Zo zou men kunnen stellen dat een dergelijke instructie leerlingen ervan afbrengt een interne 'locus of control' te ontwikkelen, omdat daarvoor de invloed van beoordeelden op het oordeel noodzakelijk is. We kozen ervoor in de instructies voor leerlingen op te nemen dat diversiteit in oordelen kan voorkomen, waaraan dat kan liggen en hoe er mee omgegaan kan worden. Verder verwachten we van de zelfbeoordelingsopdracht een gunstig effect: de relatie tussen het eigen oordeel en andermans oordelen was voor de schrijver duidelijk.

Om de acceptabiliteit van oordelen te verhogen, nemen we de suggestie van Cederblom en Lounsbury (1980) over de beoordeling formatief te laten zijn. Leerlingen moeten nadat ze feedback krijgen in staat gesteld worden iets met de feedback te doen. Ze moeten het opstel kunnen herschrijven. Daarbij is het minder wenselijk de leerlingen te verplichten de voorgestelde revisies te effectueren. De verantwoordelijkheid voor het reviseren kan beter bij de leerlingen zelf liggen.

Processtudies. Het is aannemelijk dat leerlingen van 15 jaar, onze doelgroep, in staat zijn teksten te becommentariëren. Zij lijken in twee rollen te reageren op teksten: als natuurlijke lezers, zeker als de tekstinhoud samenvalt met hun referentiekader, en als (beginnend) tekstcriticus (de instructierol). Beide rollen zouden zij moeten kunnen vervullen in een didactiek van leerlingenrespons, om verwarring van rollen te voorkomen. In de experimentele onderwijsprocedure zullen leerlingen eerst lezen en reageren en daarna pas commentaar geven. De rol van instructeur moet ontwikkeld worden. Deze ontwikkeling heeft tijd nodig en moet geleid worden via het aanbieden van tekstcriteria. Door de volgorde van de tekstcriteria vast te leggen, kan bereikt worden dat lezers niet alleen op oppervlaktekenmerken letten. Het is voor schrijvers beter als zij het commentaar schriftelijk krijgen. Leerlingenrespons moet alleen gebruikt worden als het bij de schrijfopdracht past: anders beschouwen schrijvers medeleerlingen niet als valide commentatoren.

Leerlingen lijken geholpen te moeten worden de feedback te verwerken: zwakkere leerlingen zouden anders alleen op oppervlakteniveau reviseren. Leerlingen gaan langzamerhand de tekstcriteria op eigen teksten toepassen, maar moeten geholpen worden de teksten te verbeteren. Een mogelijkheid om de

revisietaak te structureren, is het opstellen van een herschrijfplan. Het moet leerlingen duidelijk zijn dat zij zèlf bepalen welke keuze zij maken uit de feedback, opdat zij niet te afhankelijk reageren op andermans oordelen en de zeggenschap over de eigen tekst verliezen. Er lijkt heel wat tijd gemoeid met de transfer van het kritisch inzicht in de feilen van de eigen teksten, naar het verbeteren van die teksten. Een onderwijsprogramma kan daarom niet kort zijn.

Effectstudies. De belangrijkste les die we uit deze studies trokken was die van het onderscheidingsvermogen van de statistische toetsen en de klakkeloze toepassing van statistische toetsen bij niet onafhankelijke waarnemingen. Het onderscheidingsvermogen dient volgens conventie 80% te zijn; in slechts enkele studies bereikt men dit. Soms zelfs ten onrechte: als men zou corrigeren voor de intraclasscorrelatie, zou dit niet het geval zijn. Dit gegeven bracht ons tot de volgende strategie. Om docenteffecten vast te kunnen stellen moet elke docent zowel een experimentele als een controlegroep lesgeven. Een klas wordt gezien als de onderzoekseenheid. Via loting wordt een van beide klassen aan een conditie toegewezen. Het resulterende onderzoeksontwerp is dat van een echt experiment. De analyses zullen dan echter ook op klasseniveau verricht dienen te worden. Gesteld dat we erin slagen tien docenten bij ons onderzoek te betrekken, dan zou dit een onderscheidingsvermogen van 18% betekenen (bij een middelmatige effectgrootte bij $\alpha$=.05). Een dergelijk experiment is niet verstandig, maar een uitbreiding van het aantal observatie-eenheden (klassen) tot 62 per conditie om het gewenste onderscheidingsvermogen te halen, is onmogelijk. Wat we wel kunnen doen, is: in tweede instantie analyseren op leerlingenniveau, met een correctie van het aantal vrijheidsgraden. Bij $\alpha$=.05, een effectgrootte van .25 en een onderscheidingsvermogen van 80%, zouden we 62 leerlingen nodig hebben (ongecorrigeerd) in elke conditie. Afhankelijk van de intraclasscorrelatie kan geschat worden hoeveel proefpersonen we extra nodig zouden hebben. Het streven was dus zoveel docenten bij het onderzoek te betrekken als mogelijk en vervolgens steekproeven uit elke klas te trekken, zodanig dat het effectieve aantal proefpersonen groot genoeg was (zie Kish, 1965, p.162).

In de effectstudies bleek niet één afhankelijke variabele veelbelovend. Veelal was men gericht op formele kenmerken. Uit de processtudies blijkt dat men inderdaad op dat niveau verbeteringen aanbrengt. Onze verwachting dat vooral retorische kenmerken van de tekst, doel- en publiekgerichtheid, onder invloed van leerlingenrespons zouden verbeteren, werd noch bevestigd, noch tegengesproken door de resultaten van deze studies.

Wat de onafhankelijke variabele betreft, het gegeven stelonderwijs, stelden veel onderzoeken teleur. Er was slechts een beperkt aantal studies waarin de experimentele en controleconditie alleen in de feedbackfase verschilden. Wij werden

versterkt in de gedachte de beide condities zo weinig mogelijk van elkaar te laten verschillen. Of er binnen de variaties van leerlingenrespons meer of minder werkzame waren, kon uit een vergelijking van de succesvolle en onbesliste studies niet opgemaakt worden.

# 4. SAMENSTELLING VAN DE EXPERIMENTELE ONDERWIJSPROCEDURE

Om te onderzoeken of het inschakelen van leerlingen als commentatoren van elkaars teksten effecten op aspecten van stelvaardigheid teweeg zou brengen als hun prestaties vergeleken werden met die van leerlingen, wier teksten door docenten becommentarieerd werden, moest een onderwijsprocedure ontworpen worden die zowel in de experimentele als in de controlegroepen gebruikt zou kunnen worden. Ten behoeve hiervan werden enkele vooronderzoeken verricht. In hoofdstuk 3 vatten wij aanbevelingen uit literatuur over leerlingenrespons samen. We analyseerden tevens een tiental lespraktijken in Nederland en verzamelden ervaringen van leerlingen die vertrouwd waren met leerlingenrespons (par. 4.1.). Op basis van de literatuur en de praktijkinventarisatie werd een eerste versie van de experimentele onderwijsprocedure samengesteld (par. 4.2.). Deze versie werd een jaar lang beproefd in vier derde klassen (H.A.V.O. en V.W.O.). Tijdens dit cursusjaar werd een vijftal bruikbaarheidsonderzoekjes verricht, ten einde de onderwijsprocedure te optimaliseren (par. 4.3.). De beproevingen leidden tot een herziene procedure. Deze wordt beschreven in de slotparagraaf van dit hoofdstuk (par. 4.4.).

## 4.1. Praktijkinventarisatie

De ondernomen praktijkinventarisatie had tot doel meer te weten te komen over werkwijzen en ervaringen van leraren die op de een of andere wijze leerlingen inschakelen in het onderwijsleerproces van het stelonderwijs. Daarnaast wilden we van leerlingen weten hoe zij de werkwijze ervoeren. Beide groepen van informanten zouden ons informatie kunnen verstrekken, op basis waarvan we een optimale concretisering van leerlingenrespons zouden kunnen realiseren.

Docenten werden individueel, leerlingen groepsgewijs geïnterviewd. Aan leerlingen werd vooraf bovendien een vragenlijst met open vragen voorgelegd. Via een enquête, verricht in het kader van een nationale studie over het stelvaardigheidsonderwijs, onder auspiciën van International Association for the Evaluation of Educational Achievement (Damhuis, De Glopper en Wesdorp, 1983) konden wij 57 scholen achterhalen die 'enigszins tot veel' gebruik maakten van leerlingenrespons. Elf leraren bleken na een uitnodigend schrijven bereid tot een interview over hun motieven en werkwijzen. Van zeven van hen interviewden we groepjes leerlingen, alsmede twee groepjes leerlingen die reeds de experimentele leergang volgden op een proefschool. In totaal werden 48 leerlingen geïnterviewd. Zij, en twee tijdens de interviews verhinderde leerlingen, vulden ook de vragenlijst in. De resultaten van dit deelonderzoek zijn gerapporteerd in Triesscheijn, Bochardt en Rijlaarsdam (1984).

De lespraktijken op tien scholen werden op een aantal aspecten beschreven (zie bijlage 3). Het overzicht riep bij ons vragen op over de opdrachtsituatie en over de commentaarfase. De opdrachtsituatie vonden wij niet helder geformuleerd: er werd geen publiek in vermeld, er werd meteen een definitieve versie gevraagd, er werd niet altijd vermeld of er een commentaarfase zou volgen. Uit het gegeven dat docenten vrijwel altijd definitieve versies lieten schrijven, concludeerden we, dat zij het leereffect van lezen en becommentariëren voor de lezer benadrukten. De verwachting was blijkbaar dat deze het geleerde in een volgende taak zou toepassen. De schrijvers konden met het commentaar niet veel meer doen dan het geleerde in een volgende taak toepassen, die hen waarschijnlijk voor weer nieuwe problemen stelde. De praktijken toonden ons dat in de commentaarfase weinig aandacht aan het commentaar geven en verwerken werd besteed. Gezien de literatuur over commentaarprocessen (hoofdstuk 3) leek het ons verstandig leerlingen hulp te bieden via fasering en hulpopdrachten om deze taken goed uit te kunnen voeren.

Uit de interviews met leerlingen en hun antwoorden in de vragenlijst, leerden we hoe zij leerlingenrespons ervoeren. Heel kort samengevat bleek ons het volgende. Als leerlingen weten dat medeleerlingen hun opstel zullen lezen, zullen ze wat terughoudender zijn met informatie over persoonlijke ervaringen en zullen ze de inhoud wat luchtiger presenteren. Het schrijven gaat hun gemakkelijker af, omdat de lezers minder streng zijn dan de docent. Het lezen van elkaars opstellen vinden ze leuk en leerzaam: ze doen ideeën op, herkennen fouten en ontwikkelen een beeld van wat een goed opstel is. Leerlingen die onzeker zijn over de kwaliteit van hun opstel voelen zich niet prettig op het moment dat hun opstel gelezen wordt.

De fasen van commentaar geven en ontvangen zijn problematischer. Leerlingen rapporteren, deels bekende (hoofdstuk 3), weerstanden. Leerlingen vinden het lastig negatief commentaar te geven: ze vinden het vervelend te kritiseren, zitten met de vraag of de geponeerde kritiek terecht is en ervaren een gebrek aan strategieën om negatief commentaar te formuleren zonder de geadresseerde te kwetsen. Een beoordelingsschema wordt als steun gezien bij de uitvoering van deze taak. Ook blijken affectieve relaties leerlingen parten te spelen bij het geven van vooral negatief commentaar. Leerlingen zijn bang 'teruggepakt' te worden of voor 'verwaand' te worden versleten. Leerlingen twijfelen eraan of de commentaartaak met voldoende inzet en strikt eerlijk wordt uitgevoerd. Op het accepteren van het commentaar hebben dergelijke denkbeelden een negatief effect. Leerlingen blijken wel heel nieuwsgierig te zijn naar het commentaar van hun medeleerlingen. De eigen schatting van de kwaliteit van het opstel speelt in deze fase, evenals in de leesfase, een rol: als leerlingen zeker zijn van zichzelf, dan voelen ze zich prettig. Leerlingen voelen zich bij het ontvangen van commentaar nogal eens onprettig. Het commentaar valt anders uit dan leerlingen verwachten, ze zijn het niet eens met het commentaar etc.

Al vinden ze commentaar geven niet leuk, ze zien wel in dat er geleerd wordt. Leerlingen blijken vrij algemeen positief te oordelen over het nut van deze taak: het is goed andere leerlingen te helpen en je wordt er ook wat wijzer van. Je leert fouten te herkennen en kritischer het eigen opstel lezen. Je leert bovendien wat goede opstellen zijn, zodat je duidelijker weet waar je naartoe moet werken. Een mening leren vormen over opstellen wordt ook als leereffect gemeld. Ook van het commentaar ontvangen worden leereffecten onderkend: schrijvers komen te weten wat lezers over het opstel denken, krijgen aanwijzingen voor verbeteringen en denken een volgende keer op bepaalde punten te zullen letten.

Leerlingen willen het liefst commentaar op de inhoud geven. Zij achten hun medeleerlingen vooral inzetbaar als commentatoren als het de inhoud van het opstel betreft, docenten zouden vooral de niet-inhoudelijke aspecten voor hun rekening moeten nemen zoals taalgebruik en structuur. Als leerlingen echter gevraagd wordt kenmerken van een goede tekst te noemen, overwegen de meer technische aspecten zoals opbouw en taalgebruik, boven de inhoudelijke aspecten. Dit kan duiden op een sociaal wenselijk antwoord, maar is ook een aanwijzing dat opstellen voor leerlingen op twee wijzen functioneren: op een natuurlijke en op een schoolse wijze. Blijkbaar willen zij van hun docent leren, en vervullen medeleerlingen een natuurlijke behoefte dat er kennis genomen wordt van de ontwikkelde ideeën in de tekst. Commentaar van medeleerlingen wordt niet voetstoots geaccepteerd, dat van docenten wel. Zij zien de docent als een sterke representant van de norm: zijn goed/fout opmerkingen worden geaccepteerd, omdat hij daarvoor opgeleid is en meer ervaring heeft op dat gebied. Opvallend is dat leerlingen aangeven graag commentaar op de inhoud te hebben, maar er tevens blijk van geven juist dit commentaar niet te accepteren. Ze aanvaarden wel commentaar op 'zakelijke' aspecten, zoals spelling. Zodra er dus geen sterke normen voor goed en fout bestaan, laten schrijvers het commentaar voor wat het is. Leerlingen zeggen vooral kritiek naast zich neer te leggen als er alleen een oordeel gegeven wordt. Schrijvers willen weten hoe het komt dat een lezer negatief oordeelt, zij willen van de lezer meer informatie over het waarom.

Conclusies. De resultaten van de praktijkinventarisatie en de literatuurverkenning leidden tot het ontwerp van een eerste versie van de experimentele onderwijsprocedure die in de volgende paragraaf beschreven wordt. Uit de praktijkinventarisatie concludeerden we het volgende.

De opdrachtsituatie. Opstelopdrachten die een persoonlijke tekst vragen, kunnen leerlingen in moeilijkheden brengen als zij weten dat medeleerlingen hun tekst zullen lezen. Zakelijke opdrachten lijken daarom meer geschikt. Leerlingen moeten heel goed weten hoe hun tekst verder zal functioneren. Als zij weten wie de tekst zal lezen, - de docent of medeleerlingen -, kunnen zij hun keuzen qua inhoud en presentatie daarop baseren.

Lezen. De leesfase levert alleen voor leerlingen met een gering vertrouwen in eigen kunnen problemen op. Met twee maatregelen zouden ook die leerlingen wellicht meer op hun gemak gesteld kunnen worden: a) accentueren dat lezen (en becommentariëren) van opstellen bedoeld is om elkaar te helpen, om het opstel te verbeteren voordat het ingeleverd wordt; b) accentueren dat de meeste leerlingen fouten maken: voor sommige leerlingen bleek dit inzicht ertoe te leiden dat zij commentaar en lezersreacties met open vizier tegemoet zagen.

Commentaar geven. Leerlingen vinden het nuttig om te doen, maar niet leuk. De motivatie de taak zo goed mogelijk uit te voeren zal op verschillende manieren worden bewerkstelligd. a) Door een herschrijffase op te nemen in de opdrachtenreeks wordt duidelijk dat goed commentaar een ander helpt een betere versie te schrijven. Er kan dan een beroep gedaan worden op solidariteit. Na de herschrijffase worden de teksten nogmaals gelezen, commentatoren kunnen dan ervaren of hun bemoeienis resultaat heeft gehad. Daarvan kan een bemoedigende werking uitgaan. b) Door een reflectietaak toe te voegen na de commentaartaak, die leerlingen opdraagt na te gaan wat zij van die taak geleerd hebben, wordt de leerzaamheid van die taak benadrukt. c) Door anonimiteit na te streven worden storende invloeden van affectieve relaties tussen schrijvers en commentatoren ingedamd. d) Door leerlingen in staat te stellen inhoudelijk commentaar te geven wordt voldaan aan hun natuurlijke behoefte; in een commentaarschema zullen de eerste vragen betrekking hebben op inhoud. e) Door te wijzen op de validiteit van elk serieus oordeel en op het ontbreken van vaststaande normen voor een goede tekst, zijn leerlingen met weinig vertrouwen in hun oordeelvaardigheid wellicht geholpen. Door in het commentaarschema de commentatoren te helpen hun kritiek te formuleren, worden leerlingen die moeite hebben met het formuleren van kritiek gesteund.

Commentaar ontvangen en verwerken. Vier maatregelen werden getroffen om in problemen van leerlingen in deze fase te voorzien. a) Door de docent op gezette tijden het commentaarwerk van leerlingen te laten bekijken op volledigheid, - zijn alle aspecten aan bod gekomen -, zou leerlingen enigszins gegarandeerd kunnen worden dat er met inzet is gewerkt. Vermeden moet echter worden dat de commentatoren aan 'image building' gaan doen (par. 3.1.). b) Door in de commentaartaak ook opdrachten op te nemen waaruit het de schrijver kan blijken dat er aandacht aan besteed is, kan de schrijver met meer zekerheid de waarde van het commentaar schatten. c) Door een opdracht te formuleren waarin schrijvers zelf hun opstel waarderen, voordat de tekst door anderen zou worden gelezen, zouden ze zich minder onzeker hoeven te voelen. d) Door commentatoren en schrijvers over hetzelfde onderwerp te laten schrijven, zou inhoudelijk commentaar aan autoriteit kunnen winnen.

## 4.2. De experimentele onderwijsprocedure: eerste versie

De conclusies uit het literatuuronderzoek (hoofdstuk 3) en de praktijkinventarisatie vormden het fundament voor de ontwikkeling van de experimentele onderwijsprocedure. Alleen de procedure met leerlingenrespons werd ontwikkeld. Bij de ontwikkeling werd erop toegezien dat dezelfde procedure gebruikt kon worden in klassen waarin docentfeedback gegeven zou worden. De verschillen tussen de onderwijsprocedure in de experimentele en controleconditie worden in par. 4.4. getoond. De leergang werd ontwikkeld en beproefd gedurende het cursusjaar 1983-1984 op een school waaraan de onderzoeker deeltijds verbonden was. Met twee collega's ontwierp en besprak hij de procedure en het lesmateriaal. Voor een deel berustte het lesmateriaal op een op die school gebruikte syllabus voor het vakonderdeel schrijven. Verder is gebruik gemaakt van Bruffee (1980) en Rijlaarsdam en Blok (1981). Uitgegaan werd van een cursusduur van één cursusjaar, omdat leerlingen de tijd moesten krijgen de ontwikkelde kritische leeshouding om te zetten in revisiestrategieën. In het onderstaande wordt de onderwijsprocedure en het lesmateriaal kort beschreven.

Schrijfopdrachten werden gekozen uit de verzameling H.A.V.O.-A examenopdrachten. Deze opdrachten voldeden aan onze eis van zakelijkheid. Het zijn opdrachten waarin een probleemstellende alinea gevolgd wordt door een stuk of zes citaten die verschillende argumenten en zienswijzen bevatten, met een bronvermelding. Die opdrachten werden gekozen, die relatief vaak door examenkandidaten werden gekozen. De opdrachten werden zó herschreven, dat de schrijftaak communicatiever werd. Expliciet werden het doel (overtuigen), het publiek (medeleerlingen), het medium en de aanleiding vermeld. Leerlingen wisten dat hun opstel becommentarieerd en herschreven zou worden. Alle leerlingen schreven over dezelfde opdracht, opdat zij elkaar inhoudsuggesties konden geven en de modelwerking optimaal zou zijn. De leerlingen hadden een week de tijd de opdracht te bestuderen en aanvullende informatie te raadplegen. De opstellen werden op school geschreven.

Na de voltooiing van de eerste versie van het opstel, kreeg de leerling de gelegenheid een eigen oordeel over het opstel te formuleren. De docent kopieerde alle, gecodeerde, anonieme opstellen driemaal en verdeelde de kopieën willekeurig over klasgenoten.

Dan volgde de commentaarfase. Er werd schriftelijk en anoniem commentaar gegeven. We onderscheidden twee fasen: een lees- en reageerfase en een commentaarfase. Uit de literatuur en de praktijkinventarisatie was het nuttig gebleken onderscheid te maken tussen de rol van lezer en die van commentator. Daarmee sloten we aan bij Elbow (1981, p.251) die twee soorten feedback naast elkaar plaatst: 'reader based feedback' en 'criterion based feedback'. Door beide typen feedback van leerlingen te vragen, is een synthese tussen de opvattingen

Tabel 4.1.: Fragmenten uit twee feedbackinstrumenten (overgenomen uit Rijlaarsdam, Groeneweg en Weijers, 1983)

Tabel 4.1.A.: Fragment uit het subjectieve reactieformulier.

F.

| | zeer mee eens | | | | zeer mee oneens |
|---|---|---|---|---|---|
| 3. Ik werd echt bij het onderwerp betrokken | 1 | 2 | 3 | 4 | 5 |
| 4. In de inleiding kreeg ik het gevoel dit moet ik lezen | 1 | 2 | 3 | 4 | 5 |

Tabel 4.1.B.: Fragment uit het objectiverende commentaarformulier.

Fragment uit het objectiverende reactieformulier: publiekgerichtheid

6. Wordt de lezer in de inleiding tot verder lezen aangespoord en bevat het slot een uitsmijter?

| 1 Nee 2 | 3 Ja | 4 Ja 5 |
|---|---|---|
| Het begin is te traag, de lezer haakt bijna af. Het begin is te gewoon, voorspelbaar. De inleiding spreekt niet aan. Uitsmijter ontbreekt | De aandacht van de lezer wordt zo wel getrokken. Je kunt nog wel iets aan het begin verbeteren om nog meer succes te hebben. Uitsmijter kan beter | Dit is een begin waardoor je de aandacht van de lezer trekt. Zo wordt de lezer geraakt. De inleiding spreekt aan. Prima uitsmijter |

7. Wordt de lezer het gevoel gegeven dat hij echt met het onderwerp te maken heeft?

| 1 Nee 2 | 3 Ja | 4 Ja 5 |
|---|---|---|
| De lezer zal zich niet betrokken voelen bij het onderwerp. De lezer valt in slaap | Je doet je best (zie mijn 7 in de kantlijn) om de lezer te laten voelen dat hij iets met het onderwerp te maken heeft. Helaas blijft het bij een poging. Probeer er meer aandacht aan dit aspect te besteden bij herschrijving | Je roept (zie mijn 7 in de kantlijn) gevoelens, ervaringen bij de lezer op waardoor deze zich betrokken zal gaan voelen |

8. Blijkt in de tekst dat de schrijver beseft dat er lezers zijn en wie zijn lezers zijn?

| 1 Nee 2 | 3 Ja | 4 Ja 5 |
|---|---|---|
| Noch uit de manier van schrijven, noch aan de informatie is te merken dat je schrijft voor lezers van ± 15 jaar | Maar je zou (aanstrepen!)<br>o *de manier van schrijven*<br>o *de informatie*<br>nog meer op deze lezers toe kunnen spitsen. Er zitten wat zaken in die niet zo interessant zijn voor dit publiek (zie 8 in de kantlijn) | Dat blijkt duidelijk uit *de manier van schrijven* (spreekt bijv. de lezer toe) *en de informatie die hij geeft*: de voorbeelden, anecdoten, verduidelijkingen hebben allemaal te maken met deze lezersgroep. Hij roept gevoelens op die juist van toepassing zijn op de lezers |

9. Is het opstel makkelijk te lezen voor dit publiek?

| 1 Nee 2 | 3 Ja | 4 Ja 5 |
|---|---|---|
| Lezers zullen grote moeite hebben er iets van te begrijpen | Soms zal de lezer even moeilijk kijken. Door verbeteringen in: (aankruisen)<br>o presentatie<br>o formulering (zinnen, woordkeus)<br>o inhoud (toevoegingen)<br>Op de plaatsen met een 9 in de tekst, kun je je tekst nog wel begrijpelijker maken | De informatie is duidelijk, wordt helder gepresenteerd en geformuleerd |

10. Legt de schrijver alles heel duidelijk uit?

| 1 Nee 2 | 3 Ja | 4 Ja 5 |
|---|---|---|
| De informatie is warrig gepresenteerd, en je helpt de lezer niet iets te begrijpen. Het lijkt alsof je denkt dat de lezer alles al weet | In de meeste gevallen. Alleen bij 10 in de kantlijn neem je te grote stappen. Je zou daar nog wel iets kunnen verbeteren | Je doet uitstekend je best om zaken uit te leggen waarvan je denkt dat je lezers daar iets minder over weten dan jij. Stap voor stapje ga je op je doel af |

van Bruffee (1980), Elbow (1974) en Moffett (1968) tot stand gebracht: de leerling is lezer, deelnemer aan de communicatie, en coach, beschouwer van de communicatie. Voor beide rollen ontwierpen we een feedbackformulier. Basis voor de constructie van beide formulieren waren de items van een bestaand en reeds geëvalueerd leerlingenbeoordelingsinstrument (Blok en Rijlaarsdam, 1981; Rijlaarsdam en Blok, 1981). Aan dit instrument lagen zes categorieën ten grondslag: doelgerichtheid, publiekgerichtheid, structuur, betrouwbaarheid, nieuwswaarde en taalgebruik. Deze categorieën werden in deze volgorde opgenomen in de nieuwe instrumenten, het zg. reactieformulier en het zg. commentaarformulier. In het reactieformulier werden per categorie twee stellingen geformuleerd na selectie van items die een hoge item-rest-correlatie vertoonden in het 1981-instrument. Op de stellingen konden lezers reageren door aan te geven in welke mate ze het eens waren met deze stelling. Tevens moesten zij in enkele woorden de indruk noteren die het opstel op hen maakte.

In het tweede formulier, het commentaarformulier, werden per categorie vijf commentaarvragen opgenomen, ook op basis van item-rest-correlaties in het oude instrument. In deze vragen werd niet iets over de lezer gevraagd, maar iets over kwaliteiten van de tekst. Leerlingen konden op een vijfpuntsschaal noteren in welke mate hun antwoord bevestigend dan wel ontkennend was. Om de inhoud van de vraag en de schaalpunten te verduidelijken èn om de commentaargever tot explicitering van zijn antwoord te brengen, werd onder de schaalpunten 1, 3 en 5 een tekstje geschreven (zie tabel 4.1.). Commentatoren konden bij veel vragen niet volstaan door middel van een cijfer antwoord te geven en schaalwaarden aan de schrijver te rapporteren, maar moesten vaak een toelichting geven of het nummer van de vraag noteren in het opstel, zodat de schrijver kon nagaan op welke passage het commentaar betrekking had (zie tabel 4.1., vraag 7, schaalpunt 4 en 5). Dit zou de acceptabiliteit van de feedback verhogen. Bovendien werd van de commentator gevraagd het doel van de tekst te formuleren, publiekgerichte zinnen te onderstrepen, hoofdzinnen van alinea's te onderstrepen en de overgang tussen inleiding en middenstuk en slot en middenstuk te markeren. Met deze extra taken werd beoogd dat de commentator de tekst intensief zou lezen en dat de schrijver over aanwijzingen beschikte in welke mate de commentator werk gemaakt had van zijn taak. Verondersteld werd dat de schrijver via dit formulier over atomistische tekstgerichte feedback beschikte, waarin zowel commentaar op vorm als inhoud was opgenomen. Om leerlingen in de gelegenheid te stellen meer informatie over de betekenis van de zes categorieën te leren, werd een achtergrondtekst van enkele pagina's geschreven over de zes categorieën. Deze tekst konden leerlingen ook raadplegen als zij feedback verwerkten en revisies bedachten.

Om de commentaartaak aantrekkelijker te maken voor de leerlingen, werd een beroep gedaan op de nuttigheidsfunctie

die leerlingen in de praktijkinventarisatie onderkenden. Na het lezen en na het becommentariëren vervulden de leerlingen schriftelijk een reflectie-opdracht, waarin gevraagd werd de zelfwaargenomen leereffecten te noteren.

De leerlingen ontvingen vervolgens via de docent hun drie anoniem becommentarieerde opstellen terug: drie kopieën met aantekeningen en drie of meer multobladen waarop itemnummers, scores en toelichtingen geschreven waren. Om de leerlingen te helpen de grote hoeveelheid feedback te verwerken, werd deze taak in stappen verdeeld: ordenen, samenvatten, kiezen, opstellen van herschrijfplan. In de instructie werd benadrukt dat leerlingen zelf verantwoordelijk waren voor hun keuze uit alle feedback en het recht hadden feedback naast zich neer te leggen. Wel werd geeist dat leerlingen hun keuzen zouden verantwoorden. Na het opstellen van het herschrijfplan werd gedurende één lesuur de definitieve versie herschreven. Een paar lesuren later rouleerden de mappen door de klas om elkaars laatste versies te kunnen lezen.

De experimentele onderwijsprocedure werd ingepast in het bestaande lessenplan. In de eerste twee trimesters zouden leerlingen twee van deze schrijfronden van 6 à 10 lesuren voltooien, in het laatste trimester één. De multomap die alle opdrachten bevatte, werd na elke schrijfronde per trimester ingeleverd en kwantitatief beoordeeld als onderdeel van het rapport. Tevens kozen leerlingen een van de twee trimester-opstellen uit, en legden die aan de docent voor een rapport-cijfer voor. De docent beoordeelde het commentaar dat leerlingen gaven niet inhoudelijk, om een 'image building' effect tegen te gaan. Er werd slechts nagegaan of de taak verricht was.

Gedurende het cursusjaar werd de experimentele onderwijsprocedure beproefd op verschillende wijzen en momenten. Daarvan doen we verslag in de volgende paragraaf. Sommige resultaten van de beproevingen leidden tot revisies.

## 4.3. Vijf bruikbaarheidsonderzoeken

In de loop van het cursusjaar waarin de experimentele onderwijsprocedure werd uitgeprobeerd door twee docenten en de onderzoeker in vier derde klassen, zijn een vijftal bruikbaarheidsonderzoeken verricht om informatie te verzamelen over de wijze waarop de onderwijsprocedure functioneerde. Op basis van de resultaten van deze onderzoeken brachten we veranderingen aan in het leerlingenmateriaal en de procedures. We vatten de opzet en de resultaten van deze onderzoekjes kort samen. In de volgende paragraaf (4.4.) relateren we beslissingen aan wijzigingen in de experimentele onderwijsprocedure aan deze onderzoeken.

Beproeving 1: procesonderzoek. Doel van dit onderzoek was na te gaan welke problemen leerlingen op de proefschool ervoeren

bij het werken met de twee feedbackinstrumenten, - het reactieformulier en het commentaarformulier -, in hun rol van commentator en in de rol van schrijver die het commentaar verwerkt. Negen Atheneum-3 leerlingen uit drie verschillende klassen, die eenmaal de gehele lescyclus in klasseverband hadden gevolgd, werden voor de tweede cyclus uit het klasseverband gehaald en voerden enige taken hardopdenkend uit. De protocolanalyse is gerapporteerd in Bochardt en Rijlaarsdam (1984).

Commentatoren. De interpretatie van de in de formulieren opgenomen criteria verschilde soms van wat bedoeld werd, waardoor de validiteit van de formulieren in het gebruik inboet. Ook bleek dat leerlingen niet aan alle vijf vragen per categorie evenveel waarde hechtten: er vond een niet in de instructie opgenomen weging plaats. Wel was het zo, dat leerlingen ongeveer dezelfde vragen als de kern van een categorie beschouwden.

Naast dit inzicht over de betekenis van de commentaarvragen voor leerlingen, constateerden we een vijftal problemen voor commentatoren. Het eerste probleem betrof de vragen in de formulieren. Leerlingen moesten teveel vragen beantwoorden, hadden problemen met de interpretatie van sommige vragen en ervoeren overlap tussen de vragen. Het tweede probleem betrof het geven van een waardering. Soms konden leerlingen geen waardering toekennen, soms wilden ze het niet. Ze konden het niet als het bevraagde tekstelement in het opstel ontbrak of niet duidelijk te identificeren was. Het niet willen waarderen sproot voort uit het geringe belang dat leerlingen hechtten aan sommige tekstelementen. Vooral vragen in de categorieën betrouwbaarheid en taalgebruik leken van weinig belang. De gegeven situatie dat alle leerlingen dezelfde opstelopdracht hadden gekregen, was er ook weleens de oorzaak van dat commentatoren in de problemen geraakten. Vooral vragen in de categorie nieuwswaarde ('een nieuwe kijk op het onderwerp?') en betrouwbaarheid waren voor sommige leerlingen daardoor moeilijk te beantwoorden. Een derde probleem voor commentatoren was het gebruik van de vijfpuntsschaal. Er deden zich problemen voor met betrekking tot schaalpuntomkeringen, de keuze voor een schaalpunt en de toelichtingen bij schaalpunten. De schaalpuntomkeringen hadden alle betrekking op die stellingen in het reactieformulier, die een enkele of zelfs een dubbele ontkenning bevatten. In het commentaarformulier hadden commentatoren meermalen moeite met het kiezen van een schaalpunt, omdat zij zich niet konden verenigen met de omschrijving die bij het schaalpunt vermeld werd.

Naast deze problemen met de didactische instrumenten signaleerden we twee algemenere problemen: motivatieproblemen en rolproblemen. Op een gegeven moment verlichtten de commentatoren hun taak door zich te beperken tot het geven van scores en alle andere verplichtingen te verzaken, zoals toelichtingen, vraagnummers in het opstel, het onderstrepen van publiekgerichte zinnen en het formuleren van de doelstelling van de tekst.

Meer dan de helft van de commentatoren had moeite met de rol die zij vervulde. 'Voor mij is het opstel de moeite waard, maar voor lezers?' 'Als ik het publiek ben, dan...' Zijn leerlingen nu lezers, het publiek of moeten zij andermans reacties schatten? Wat moet de referentie voor het oordeel zijn dat zij vellen? Dit rolprobleem deed zich voor bij vragen uit de categorie nieuwswaarde. Een ander rolprobleem trad op bij vragen uit de categorie taalgebruik. Leerlingen voelden zich niet capabel hierover uitspraken te doen: 'Ik heb er niet veel verstand van, daar moet een leraar maar op letten.'

Herschrijven. Hoewel leerlingen worstelen met de enorme hoeveelheid feedbackgegevens, hun ideeën over de kwaliteit van hun opstel en soms het gebrek aan overeenstemming tussen commentatoren, kiezen ze toch vrij intuïtief die categorieën van de feedback, waarop zij relatief het meeste commentaar kregen. We verzamelden per leerling de frequenties van de scores die betekenen: verbeter! (scores 1,2,3). De ratio van de frequentie van 'verbeterscores' en 'houden zo scores' definieerden we als een verbeterindex. Hoe hoger de index, hoe sterker het signaal voor de herschrijver dat er verbeterd moet worden. Deze indices zijn vermeld in tabel 4.2. Een asterisk duidt aan dat dit aspect voor verbetering aanvaard werd door de herschrijver in het protocol of in het herschrijfplan. Het valt op dat het aspect nieuwswaarde door deze leerlingen niet verbeterd zal worden, hoe sterk het signaal om het wel te doen ook is: van de signalen zijn er 6 sterker dan 1., wat betekent dat tenminste 50% van de commentaren zegt: verbeter! Zelfs in de heel duidelijke gevallen van Bas en Marco wordt nieuwswaarde niet verbeterd.

Tabel 4.2.: Sterkte van verbetersignalen (de ratio van frequenties van scores 1,2 en 3 en scores 4 en 5)

| | Aspect | | | | | |
|---|---|---|---|---|---|---|
| | Doelgerichtheid | Publiekgerichtheid | Structuur | Betrouwbaarheid | Nieuwswaarde | Taalgebruik |
| Richard | .7 | .6* | .6 | .7 | .8 | .8 |
| Nathalie | 2.5* | 1. | 1.7* | .8* | 2.2 | 1.3 |
| Esther | 0 | .1 | .2* | .1 | .1 | .5* |
| Simon | 1.3* | 1.3* | 2.2* | 1.3 | 3. | 4. |
| Helen | .5 | .3 | 1.2 | .8* | 1.9 | 1.* |
| Bas | 3.3 | 4.3 | .6 | 8.* | 20. | 2.* |
| Marco | 7.* | 6.* | 3. | 5. | 10. | 6. |
| Sandra | .3* | .3 | .4 | .4 | 1. | .3* |
| Saskia | .5 | .3* | 1.* | .9 | .8 | .2* |

* aspect voor verbetering aanvaard

Als we leerlingen buiten beschouwing laten die op de categorieën relatief weinig verschillend beoordeeld werden (Richard, Esther, Sandra) èn we laten de categorie nieuwswaarde buiten

beschouwing, dan blijkt dat Nathalie het sterkste signaal van de zes in haar verbeteringen betrekt evenals Bas, Marco en Saskia, en Simon de op één na sterkste, evenals Helen. Herschrijvers lichten die categorieën eruit voor verbetering, waarop zij relatief het meeste commentaar krijgen. Zij schatten de sterkte op vrij grove manier en komen tot een correct resultaat. De categorie nieuwswaarde lijkt niet te functioneren voor herschrijvers.

Uit een vergelijking tussen de opmerkingen uit de protocollen over het verbeteren op categorieën en het herschrijfplan blijkt dat slechts bij één leerling informatieverlies optrad. In een paar andere gevallen trad juist informatiewinst op: in het herschrijfplan werden zelfs meer verbeteringen beoogd dan leerlingen tijdens het verwerken en vaststellen van commentaar formuleerden. De negen leerlingen deden in totaal 19 voorstellen om hun eerste versie te verbeteren. Een viertal beoordelaars beoordeelden eerste en herschreven versies op de voorgenomen herschrijving. In negen van de 19 gevallen werd de tweede versie beter gevonden op het voorgenomen verbeteringsaspect dan de eerste versie door drie of vier van de beoordelaars. Zeven leerlingen verbeterden hun opstel zichtbaar op tenminste één aspect. In één geval bleek de eerste versie unaniem beter dan de tweede versie.

Beproeving 2: vragenlijstonderzoek. Doel van dit bruikbaarheidsonderzoek was te onderzoeken of en in welke mate de leerlingen de feedbackopdrachten uitvoerden en de gekregen feedback verwerkten en hoe zij oordeelden over het reactieformulier en het commentaarformulier. Overeenkomstig het eerste bruikbaarheidsonderzoek werd de bruikbaarheid voor twee rollen onderscheiden: de bruikbaarheid voor commentatoren en voor herschrijvers. Gedurende de vierde schrijfronde werden aan de leerlingen op de proefschool vier vragenlijsten voorgelegd. Elke vragenlijst werd verstrekt op het moment dat leerlingen aan het desbetreffende onderdeel van de experimentele onderwijsprocedure werkten. Van 88 leerlingen werden vragenlijsten retour ontvangen. De belangrijkste bevindingen worden hieronder samengevat.

Het reactieformulier. Commentatoren zeiden allen op de 12 reactiestellingen van dit formulier te reageren. De opdracht de indruk te beschrijven die het opstel maakte, werd door 80% steeds vervuld. De leerlingen waren verdeeld over de aantrekkelijkheid van deze opdracht, maar een duidelijke meerderheid vond hem zinvol (74%).

De belangrijkste oordelen over het formulier waren de volgende. Gevraagd werd naar de validiteit van het reactieformulier: toon je daarmee je eerste reactie? 'Ja' zei 74%; 'nee' vond 26%. Het invullen van het reactieformulier werd niet moeilijk gevonden: slechts 13% gaf te kennen moeite te hebben met dit formulier. Twee stellingen bleken moeilijk te zijn, beide betroffen het aspect betrouwbaarheid. Verder werd nog gevraagd naar het relatieve belang dat commentatoren hechtten aan de 12 stellingen. Verreweg de belangrijkste stelling vonden zij: 'Ik ben eigenlijk niet overtuigd', inderdaad een finaal criterium voor een betogende tekst.

Herschrijvers vonden vrijwel allen (96%) dat de 12 reactiestellingen altijd ingevuld waren. Over de opdracht een verbale omschrijving van de indruk te geven waren zij minder optimistisch dan commentatoren: 58% antwoordde ja/bijna altijd; 30% soms; 12% nee/bijna nooit.

Wat de validiteit van het formulier betreft: 38% vond dat zij via de scores te weten kwamen wat lezers van hun opstel vonden; 61% vulde hier 'matig' in. Toch zei 62% bij het herschrijven rekening te houden met de scores; 38% zei dat niet te doen. De verbale indrukomschrijving werd informatief gevonden (70%); iets meer leerlingen (76%) zeiden rekening te houden met deze omschrijving bij het herschrijven. Herschrijvers vonden het niet moeilijk om te begrijpen wat de scores betekenden (80%). Ook herschrijvers vonden 'Ik ben eigenlijk niet overtuigd' de belangrijkste stelling. Tenslotte werd nog aan herschrijvers gevraagd van hoeveel leerlingen zij een reactie wilden om goed te kunnen herschrijven: 73% antwoordde 3; 15% 2; 4% 1 en 9% gaf te kennen van meer dan drie leerlingen reacties te willen.

Het commentaarformulier. Ook hier werd eerst gevraagd naar de werkwijze van commentatoren en daarna naar hun oordelen. Commentatoren moesten eerst de hoofdgedachte formuleren: 82% zei dat te doen. Daarna moest van elke alinea de kernzin onderstreept worden: 61% zei dat te doen, 23% deed het nooit. Het markeren van de overgang tussen inleiding, kern en slot werd door 70% gedaan, 30% deed het nooit. Met een golflijntje publiekgerichte zinnen aangeven deed slechts 56%. Het beantwoorden van de commentaarvragen scoorde hoog: 95% deed dit. Bij een aantal commentaarvragen moesten leerlingen het corresponderende vraagnummer in het opstel noteren om aan te geven waar de feilen zijn. Bij 12 van de 17 vragen waarvoor dit van toepassing is zei 40% of meer dit nooit of bijna nooit te doen. Het invullen van het commentaar kostte veel tijd: 51% werkte 15 à 30 minuten, 41% 30 à 45 minuten per opstel.
Welke taken vonden commentatoren geschikt om goed commentaar te geven? Van groot belang vonden zij commentaar in het opstel zelf (98%), toelichtingen bij antwoorden op commentaarvragen (92%), verwijzingen in het opstel naar commentaarvragen (82%) en de antwoorden op de commentaarvragen (82%). Over de zinvolheid van andere taken was men verdeeld.

Het overgrote deel (94%) vond dat de schrijver door het commentaarformulier geholpen zou worden. Over de moeilijkheid van de commentaartaak en onderdelen ervan oordeelde men gedifferentieerder. Het invullen van het commentaarformulier vond 20% moeilijk; 65% antwoordde 'gaat wel' en 5% vond het makkelijk. Van de extra opdrachten werden het aangeven van publiekgerichte zinnen en het noteren van vraagnummers in het opstel moeilijk gevonden. Vrij moeilijk werd het onderstrepen van kernzinnen van alinea's gevonden. De belangrijke taken als het beantwoorden van de commentaarvragen, het toelichten van die antwoorden en het plaatsen van commentaar in het opstel zelf werden niet moeilijk gevonden. Er werd ook per commen-

taarvraag naar de moeilijkheid gevraagd. Vier vragen werden door 40% of meer van de leerlingen als moeilijk ervaren: twee vragen over doelgerichtheid, één over structuur en één over betrouwbaarheid. Als geheel bleken de categorieën doelgerichtheid en betrouwbaarheid het moeilijkst. Om de betekenis van de aspecten voor leerlingen te achterhalen, werd gevraagd welke commentaarvragen het best uitdrukten wat de commentator onder het aspect verstond. Deze gegevens konden worden gebruikt bij het selecteren en herschrijven van vragen. Voor dit doel werd ook naar het relatieve belang van de commentaarvragen gevraagd en naar de mate van duidelijkheid van de vragen. Wat dit laatste betreft vond 59% dat er wel wat vragen verduidelijkt zouden kunnen worden, waarbij het nummer van de betreffende vragen genoemd werd. Over de duidelijkheid van de toelichtende tekstjes onder de schaalpunten was men (85%) wel tevreden. Van enkele tekstjes werd gezegd dat deze wel duidelijker konden. Het aantal antwoordmogelijkheden bij de commentaarvragen vond men wel genoeg (81%); 68% vond vijf mogelijkheden precies goed, 27% zou aan drie mogelijkheden genoeg hebben. Veel leerlingen vonden het commentaarformulier te lang (42%) of veel te lang (31%). Uit de mogelijkheden om de commentaartijd te bekorten werd vooral gekozen voor het becommentariëren van minder opstellen: 32% koos voor een korter formulier en drie opstellen; 24% voor een korter formulier en minder opstellen; 44% wilde het formulier zo houden, maar minder opstellen. Gevraagd naar de minimale omvang van het formulier om een goed commentaar te kunnen geven, antwoordde bijna de helft (48%) 16-20 vragen; van 18% mocht het nog minder zijn (11-15), 15% vond 21-25 vragen een adequate lengte. Als de leerlingen voor de keuze gesteld zouden worden aspecten te moeten laten vallen, zouden ze doelgerichtheid en publiekgerichtheid willen behouden en taalgebruik en betrouwbaarheid niet.

Wat krijgen <u>herschrijvers</u> aan feedback binnen? Zij nemen de hoeveelheid feedback wat pessimistischer waar dan de commentatoren. De hoofdgedachte wordt volgens 67% van de herschrijvers geformuleerd; het onderstrepen van kernzinnen door 43%; het markeren van de overgang tussen inleiding, kern en slot door 53%; het markeren van publiekgerichte zinnen door 42%; het noteren van nummers in de kantlijn door 61%; het beantwoorden van de commentaarvragen door 98%.

Er bleken geen commentaarvragen te zijn die volgens herschrijvers vaak werden overgeslagen. Wel waren er vragen, waarbij het verplichte corresponderende nummer in de kantlijn van het opstel ontbrak: van drie vragen gaven 10 of meer leerlingen aan dat het nummer vaak ontbrak. Herschrijvers (86%) vonden de scores op het commentaarformulier behulpzaam als het ging om het verbeteren van het opstel. Een nog groter percentage (90%) gaf te kennen bij het herschrijven rekening te houden met de scores. Over de behulpzaamheid van de extra opdrachten verschilden herschrijvers nauwelijks van mening met commentatoren. Alleen het markeren van hoofdzinnen werd (nog)

minder behulpzaam geacht (26%),evenals het markeren van de overgang tussen inleiding, kern en slot (44%). Ook aan herschrijvers werd gevraagd wat de twee belangrijkste vragen per aspect waren. Hun oordelen weken niet opvallend af van die van commentatoren. Ook werd gevraagd naar de acht belangrijkste en de acht onbelangrijkste commentaarvragen. Deze gegevens zouden een rol kunnen spelen bij het selecteren en herschrijven van commentaarvragen. Over de lengte van het formulier was men verdeeld: 45% vond de hoeveelheid commentaarvragen precies goed, 51% vond de hoeveelheid te groot. Het merendeel gaf aan 20 (40%) of 25 (32%) antwoorden nodig te hebben op even zoveel vragen om goed te kunnen herschrijven. Gesteld voor de keuze de hoeveelheid commentatoren te bepalen, gaf 73% te kennen er drie nodig te hebben; 19% had genoeg aan twee. Als herschrijvers gevraagd wordt op welke twee aspecten zij in elk geval commentaar willen hebben, geven zij aan doelgerichtheid (54 x) en publiekgerichtheid (34 x) de voorkeur. Structuur, betrouwbaarheid en taalgebruik zouden wat hen betreft mogen vervallen.

Beproeving 3: statistische analyses. Het doel van dit deelonderzoek was na te gaan of de feedback die leerlingen aan elkaar geven met behulp van de twee didactische instrumenten van voldoende kwaliteit was. Tevens werd de interne structuur van de instrumenten aan een analyse onderworpen en de mate waarin elk instrument iets eigens vertegenwoordigde. Voor deze analyses werden de scores verzameld die leerlingen in de tweede schrijfronde op de proefschool toekenden. De data zijn dus onder normale condities verzameld, met alle mogelijke bronnen van ruis. Het ging ons niet om de kwaliteit van de instrumenten onder optimale omstandigheden, maar om de kwaliteit zoals die zou zijn in het gewone schoolgebruik. Elk opstel werd door drie leerlingen beoordeeld. Van 110 opstellen werden de scores verwerkt. De beoordelaarsovereenstemming werd berekend op gesommeerde scores. De overeenstemming bleek laag voor het reactieformulier (een correlatie van .50) en matig voor het commentaarformulier (.65). Het commentaarformulier bestond (par. 4.2.) uit zes categorieën van elk vijf commentaarvragen. De overeenstemming tussen beoordelaars varieerde van .52 tot .61 per categorie. De coëfficiënt voor het reactieformulier mocht laag zijn: feedbackontvangers wisten dat zij te maken hadden met eerste indrukken, die sterk lezersafhankelijk zijn. Als leerlingen echter de instructierol overnemen van de docent door hun scores op het commentaarformulier aan de schrijver terug te koppelen, is een zekere mate van overeenstemming gewenst, aangezien de ontvanger anders zeer divergente feedback zou krijgen waaruit moeilijk af te leiden zou zijn wat nu het algemene oordeel is over het opstel. Er zou naar middelen gezocht moeten worden om de overeenstemming te verbeteren. De interne consistentie van beide instrumenten bleek bevredigend (Cronbachs alpha op over beoordelaars gesommeerde scores voor het reactieformulier .83; voor het commentaarformulier .93). Ook de categorieën van het commen-

taarformulier bleken uit vrij homogene verzamelingen vragen te bestaan: Cronbachs alpha varieerde van .70 tot .83. Er bleek wel overlap tussen deze categorieën: de categoriesomscores correleerden van .51 tot .70. Enkele vragen bleken sterk te correleren met andere categorieën. Bij het selecteren en herschrijven van vragen zou met de correlaties tussen de vraag-eigen categoriecorrelaties en de vraag-andere categoriecorrelaties rekening gehouden worden. Het reactieformulier en het commentaarformulier vertoonde eveneens overlap. De correlatie tussen somscores bedroeg .71.

Beproeving 4: zelfrapportages. Het doel van dit deelonderzoek was na te gaan of leerlingen leereffecten constateerden bij zichzelf na de cursus en aan welke fase van de lessencyclus zij deze effecten toeschreven. In de praktijkinventarisatie hadden we al door leerlingen gerapporteerde relaties tussen lezen en becommentariëren van elkaars teksten en leereffecten verzameld, maar het was de vraag of het door ons geconstrueerde onderwijsprogramma eveneens in de ogen van leerlingen vruchten afwierp. De via zelfrapportages verkregen effecten zouden een rol kunnen spelen in de uiteindelijke versie van het onderwijsprogramma. Ze konden gebruikt worden als motiverende elementen. Aan het eind van het cursusjaar werd de leerlingen die de gehele cursus hadden meegemaakt, gevraagd per didactische fase te rapporteren wat zij er volgens henzelf van leerden. Uit de leerverslagen van 81 leerlingen uit vier derde klassen werden drie fragmenten geanalyseerd, namelijk de antwoorden op de vragen: 1. wat leerde ik door het lezen van andermans opstellen? 2. wat leerde ik door het becommentariëren van andere opstellen? 3. wat leerde ik door het commentaar op mijn opstellen? Per vraag werden de proposities gescoord door twee onafhankelijk scorende doctoraalstudenten, die werkten met een op basis van de praktijkinventarisatie vastgesteld categorieënschema. De gemiddelde overeenstemming tussen scoringen dezer studenten bedroeg 75% (Rijlaarsdam, 1984).

Uit de analyses bleek dat leerlingen van het lezen van elkaars opstellen over het onderwerp leerden en modellen van teksten opbouwden. Ze leerden zowel via slechte als goede voorbeelden. Een aantal uitspraken legde een direct verband tussen geconstateerde kwaliteiten in andermans opstellen, het eigen opstel en de te schrijven herziening.

Het commentaar geven had vooral tot gevolg dat leerlingen criteria verwierven en internaliseerden. Een aantal leerlingen gaf aan dat zij die toepasten als ze gingen schrijven. Anderen meldden dat zij het tijdens het becommentariëren geleerde toepasten op hun herschrijving. Een paar leerlingen schreven dat door de commentaartaak het gewone lezen was veranderd. Naast deze leereffecten die alle gemeen hebben dat er een analytischer inzicht ontstaat, gaf eenderde van de leerlingen aan iets geleerd te hebben met betrekking tot de commentaarhandeling zelf. Of men schreef dat het nu gemakkelijker was

het commentaar op de eigen tekst te begrijpen, of men rapporteerde dat de communicatieve situatie enige verplichtingen oplegde, zoals objectiviteit en behulpzaamheid.

Van het ontvangen van commentaar werd vooral geleerd wat er goed en fout aan het eigen opstel was, waardoor goed herschreven kon worden. Ook werden algemenere uitspraken gedaan, die duidelijk maakten dat leerlingen kennis over criteria verwierven, die bij sommigen een rol speelden bij het schrijven van nieuwe opstellen. Er waren ook leerlingen die schreven te hebben geleerd hoe ze zich moesten gedragen als ze commentaar kregen. Voor anderen ging van het gekregen commentaar een voorbeeldwerking uit naar het zelf commentaar geven: goede voorbeelden werden overgenomen.

Geconcludeerd werd dat alle drie de fases in de ogen van leerlingen duidelijk repercussies hadden op het schrijven. Het lezen van elkaars teksten had een eigen functie, omdat leerlingen daarover meldden dat zij over het onderwerp leerden en dat geleerde toepasten in hun herschrijvingen. Dit was een beoogd voordeel van de situatie dat schrijvers en commentatoren over hetzelfde onderwerp schreven. Overwogen zou moeten worden, ondanks bezwaren geuit in het procesonderzoek (de eerste beproeving), deze situatie te handhaven. Duidelijk werd ook dat de fase van herschrijven aan het doel beantwoordde: leerlingen gaven te kennen het tijdens de commentaarronde geleerde in de herschrijffase toe te passen.

Finale beproeving. Om na te gaan welke problemen de organisatie van een schrijfronde opleverde als verschillende docenten ingeschakeld werden die er nog vrij onbevangen tegenover stonden en om de aan het hoofdonderzoek deelnemende docenten te trainen, werd voor deze docenten een proefronde georganiseerd. Tevens werd de beoogde schrijfopdracht voor de toetsmomenten in het hoofdonderzoek uitgeprobeerd. Er werd een handleiding voor zes lessen geschreven. Het leerlingenmateriaal was zo samengesteld, dat er een grote mate van overeenstemming was met de eerste schrijfronde in het hoofdonderzoek. Leerlingen ontvingen een boekje met opdrachten, formulieren en gelinieerde pagina's. Alleen op de categorieën doel- en publiekgerichtheid werd commentaar gegeven. De vragen van de feedbackinstrumenten waren nog niet gereviseerd. Acht docenten werkten mee aan de proefronde. Uit logboeken van docenten, observaties door onderzoekers en een gemeenschappelijke evaluatievergadering met docenten werden vijf hoofdproblemen besproken en voorstellen geformuleerd.

De schrijfopdracht werd aan de moeilijke kant gevonden. Niet alle klassen hadden er moeite mee, maar enkele moesten op weg geholpen worden door de docent. Besloten werd dat de opdrachten die op de proefschool waren uitgeprobeerd nog herzien zouden moeten worden waar het de citaten betrof, die overgenomen werden uit de zogenaamde H.A.V.O.-A-opdrachten.

De periode (mei-juni) bleek bijzonder ongeschikt voor een reeks lessen. De periode wordt gekenmerkt door grote dis-

continuïteit en veel docentenbemoeienissen met examens. Besloten werd de afronding van het onderzoek begin mei 1985 te plannen.

De schrijfronde legde leerlingen een strak werkritme op. Op zich was dit geen bezwaar, maar een wat ruimere lesplanning zou voor docenten en leerlingen aangenamer zijn, zeker in het begin van het cursusjaar waarin leerlingen moesten kennismaken met deze wijze van werken. In samenhang met het in het vorige punt genoemde werd besloten niet de geplande zes, maar vier schrijfrondes in de cursusplanning op te nemen, die per stuk wat langer konden duren.

Niet in alle klassen lukte het docenten de leerlingen ervan te overtuigen dat leerlingen anoniem moesten becommentariëren. Voor het begin van het hoofdonderzoek leek het toch van belang dat docenten dit konden bereiken. Men verwachtte dat zulks in een 'verse' klas wel zou lukken. Enkele docenten deden de andere tips aan de hand, zoals het uiteenplaatsen van banken. Zeker omdat leerlingen hun papiermassa's goed moeten organiseren, was deze wenk heel plausibel voor leerlingen.

Docenten waren veel tijd kwijt met het kopiëren en sorteren van opstellen die becommentarieerd moesten worden. Vooral als absente leerlingen hun werk later inleverden of men zich moest behelpen met zeer eenvoudige kopieerapparatuur, werd het werk van docenten zeer tijdrovend en onoverzichtelijk. Besloten werd leerlingen op setjes zelf-doorschrijvend papier te laten schrijven. Met enige voorbedrukte gegevens zou dit ook het sorteerwerk van docenten aanzienlijk vergemakkelijken.

## 4.4. Definitieve experimentele onderwijsprocedure

De vijf bruikbaarheidsonderzoekjes leverden ons veel informatie over het functioneren van de experimentele onderwijsprocedure en de twee feedbackinstrumenten. We besloten alle stappen uit de eerste versie (par. 4.2.) te handhaven. We beschrijven eerst de veranderingen die we aanbrachten in het materiaal en de gehele procedure. Daarna beschrijven we het totale lesmateriaal. In hoofdstuk 6, par. 6.3., wordt een schrijfronde uitvoerig beschreven zoals die in het onderzoek werkelijk plaatsvond. Daarom blijven we hier beknopt.

Een eerste verandering was het aanbieden van een veel grotere hoeveelheid leerstof over de zes categorieën waarop de cursus gebaseerd was: doelgerichtheid, publiekgerichtheid, structuur, betrouwbaarheid, nieuwswaarde en taalgebruik. Daarmee wilden we bereiken dat de inhoud van de feedback-instrumenten een referentiekader kreeg, hetgeen de validiteit en betrouwbaarheid van de instrumenten ten goede zou komen. Herschrijvers beschikten daardoor over meer kennis om verbeteringen aan te brengen.

Een tweede verandering was het bekorten van de commentaartaak. In het proefjaar gaven leerlingen op dertig commentaarvragen antwoord. Uit verschillende beproevingen bleek dat aantal teveel. Door nu per schrijfronde twee categorieën te

introduceren, wilden we bereiken dat er een heviger concentratie op die twee categorieën zou plaatsvinden. De validiteit en de betrouwbaarheid van de feedbackinstrumenten zou daarmee gebaat zijn.

Per tweetal categorieën werd dan ook de theorie (zie hiervoor) daarover aangeboden. In de eerste schrijfronde werden doel- en publiekgerichtheid geïntroduceerd, vervolgens structuur en betrouwbaarheid en in de derde schrijfronde kwamen nieuwswaarde en taalgebruik aan bod. In de vierde schrijfronde zou een samenvattende procesgang van het schrijven van een opstel de leerstof zijn. Gezien de ervaringen in het proefjaar (zie beproeving 2) zouden commentatoren op niet meer dan twintig vragen antwoord geven. Per categorie werden vijf vragen geformuleerd, zodat in de derde schrijfronde uit de dan dertig beschikbare vragen er twintig geselecteerd moesten worden. We hielden vast aan drie commentatoren per opstel: commentatoren zouden liever minder opstellen van commentaar voorzien, maar waren als herschrijvers toch blij met drie commentatoren.

De derde verandering had betrekking op de feedbackinstrumenten. Ze werden beide gehandhaafd, maar wel gewijzigd. Op basis van correlaties, moeilijkheidsgegevens, betekenisfouten werden vragen en toelichtende tekstjes bij schaalpunten herschreven. Waar mogelijk werd in de schaalpunttoelichting gevraagd een verwijzing in het opstel te noteren. De vragen in het commentaarformulier werden stelselmatig in de derde persoon gesteld, opdat leerlingen vanuit hun coach-functie feedback zouden geven. Alle uit de beproevingen (1,2 en 3) beschikbare informatie werd gebruikt om het reactie- en het commentaarformulier te verbeteren.

Een vierde verandering betrof de begeleidende instructie bij de taken commentaar geven en commentaar verwerken. In het proefjaar hadden we vrijwel uitsluitend via mondelinge instructie toelichtingen gegeven op deze fasen en de problemen die leerlingen ontmoetten. Nu de cursus echter overgedragen moest worden naar andere docenten, was het goed op de ervaren problemen te anticiperen. Daartoe werden op basis van leerlingervaringen uit de praktijkinventarisatie en de beproevingen instructieteksten geschreven over het commentaar geven en problemen als 'wie ben ik dat ik daarover mag oordelen?' en de invloed van affectieve relaties op oordelen over teksten. Daarover werd bijvoorbeeld geschreven dat een vriend gedupeerd is als hij van de commentator vergoelijkend commentaar zou krijgen: met commentaar help je de ander een goede herschrijving te maken. Ook werden instructieteksten geschreven over het ontvangen en interpreteren van feedback. Voorgedaan werd bijvoorbeeld hoe te handelen bij een gebrek aan overeenstemming tussen commentatoren.

Een vijfde verandering gold de fase van commentaar verwerken. Leerlingen kregen een formulier waarmee zij snel inzicht konden krijgen in de grote hoeveelheid feedbackgegevens die zij kregen. Verder werd in instructieteksten voorgedaan hoe commentaar samengevat kon worden en hoe een keuze uit het commentaar verantwoord kon worden. Er werden voorbeelden en besprekingen van herschrijfplannen opgenomen.

Een zesde verandering betrof alleen de eerste schrijfronde. In de finale beproeving hadden wij voor de leerlingen een geheel voorbedrukte schrijfronde samengesteld, wat de organisatie in de klas vergemakkelijkte. Voor leerlingen die voor de eerste keer aan zo'n schrijfronde deelnamen, betekende deze voorstructurering een aanzienlijke steun in de grote hoeveelheid taken die zij moesten uitvoeren. Voor de eerste schrijfronde in het onderzoeksjaar maakten we een dergelijke voorgedrukte schrijfronde.

Een zevende verandering was van meer organisatorische aard. We stelden aan docenten zelfdoorschrijvend papier beschikbaar, waardoor zij verlost waren van het kopiëren van opstellen. Doordat elke set samengesteld was uit vier verschillende kleuren, kon er nu ook voor worden gezorgd dat elke leerling een keer als eerste werd becommentarieerd. De commentatoren zouden nu van drie leerlingen een opstel ontvangen; elk van die opstellen had een andere kleur. Door de volgorde van werken vast te leggen via een af te werken kleurvolgorde, werd van elke leerling één kopie het eerst becommentarieerd.

Een achtste verandering was het opnemen van extra-commentaartaken bij categorieën waarvoor er nog geen geformuleerd waren. De taken hebben tot doel de opstellen al in enige mate voor te bewerken vooraleer antwoorden op de commentaarvragen werden geformuleerd. Bij betrouwbaarheid moesten leerlingen in de kantlijn van het opstel aangeven (ja!, nee! of ?) in welke mate ze het met een uitspraak eens waren. Voor nieuwswaarde moest in de kantlijn (nieuw!) geschreven worden als er een element met nieuwswaarde aangetroffen werd. Voor taalgebruik moesten met commentaarvragen corresponderende nummers in de kantlijn genoteerd worden, zoals $27^-$ (te makkelijk taalgebruik), $27^+$ (te moeilijk taalgebruik), $30^-$ (onpersoonlijk taalgebruik). Bij het beantwoorden van de commentaarvragen kon niet volstaan worden met het aankruisen van een cijfer. Bij 22 van de dertig vragen moest tevens een met de vraag corresponderend nummer in de kantlijn van het opstel geplaatst worden. Bij negen van de dertig vragen moest een toelichting bij één of meer van de antwoordmogelijkheden geschreven worden.

Voor de controleconditie werd vrijwel hetzelfde instructiemateriaal gebruikt. Alleen de instructieteksten over commentaar geven en verwerken werden aangepast. De docent die de feedbacktaak vervulde in de controleconditie zou dat met dezelfde feedbackinstrumenten doen: het reactie- en het commentaarformulier. Hij zou dezelfde commentaartaken vervullen. Omdat leerlingen in de controleconditie geen feedback gaven, werden de reflectietaken 'Wat heb ik geleerd van het lezen/het becommentariëren van andermans opstellen?' vervangen door een andere: 'Wat heb ik geleerd van deze schrijfronde?' In tabel 4.3. worden de twee condities naast elkaar gezet.

De leerlingen dienden al het werk te bewaren in een multomap. Deze bevatte vier afdelingen:

Algemene informatie. Elke zending nieuw materiaal werd vergezeld door een paar pagina's informatie over hoe het pakket dit

keer in elkaar stak en opgeborgen moest worden en een routebeschrijving van de desbetreffende schrijfronde.
Informatie over schrijven en opstellen. Deze afdeling bevatte drie delen. In deel A werd de leerstof over aspecten van teksten gepresenteerd. In deel B werd ingegaan op de reflectietaken. Deel C bevatte alle informatie, waaronder de reactie- en commentaarformulieren, over het geven en verwerken van commentaar.
Werkafdeling. Voor de eerste ronde werd een geheel voorgedrukte ronde opgenomen, waardoor leerlingen begeleid werden in de organisatie van alle stappen. Verder werd deze afdeling gevuld met de schrijfopdrachten en eigen multopapier voor andere schrijfronden.
Rommelhoek. Hierin ruimde de leerling al het materiaal op dat hij produceerde tijdens een ronde, maar dat niet in de definitieve werkafdeling opgenomen moest worden.

Tabel 4.3.: Vergelijking tussen de onderwijsprocedures van de experimentele en de controleconditie

| Experimentele conditie | Controleconditie |
|---|---|
| 1. Bestuderen van instructieteksten over tekstaspecten, maken van controlevragen, bespreking in de klas, bestudering van de schrijfopdracht | 1. idem |
| 2. Uitvoeren van de schrijfopdracht, klad en net, in de klas | 2. idem |
| 3. Reflectietaak: wat vind ik van mijn opstel en hoe verliep het schrijfproces dit keer? | 3. idem |
| 4. Lezen van drie anonieme opstellen | 4. Tijdens twee lesuren waarin de experimentele conditie |
| 5. Reflectietaak: wat heb ik geleerd door het lezen van de opstellen? | 5. commentaartaken verricht, krijgt de controleconditie werk op uit de gebruikte les- |
| 6. Becommentariëren van dezelfde drie opstellen | 6. of leesmethode. De docent |
| 7. Reflectietaak: wat heb ik geleerd door het becommentariëren van de opstellen? | 7. leest en becommentarieert de opstel thuis. |
| 8. Commentaar ordenen, samenvatten, kiezen | 8. idem |
| 9. Herschrijfplan opstellen | 9. idem |
| 10. Herschrijven | 10. idem |
| | 11. Reflectieopdracht: wat heb ik geleerd van deze schrijfronde? |

Uit tabel 4.4. kan opgemaakt worden welk materiaal per schrijfronde werd toegestuurd. Het lesmateriaal werd vergezeld door een docentenhandleiding. Deze bevatte een algemeen

gedeelte en een gedetailleerde planning per schrijfronde. In het algemene gedeelte kwamen aan de orde: het verdelen van de opstellen over commentatoren, het bewaken van de werksfeer en de anonimiteit, het opvangen van absenten, het controleren van de schrijfmappen, het bijhouden van de logboeken en het beoordelen van de opstellen voor het rapport. Verder ontvingen docenten enig additioneel materiaal. Zij kregen de beschikking over doorschrijfpapier, waarop leerlingen in één schrijfhandeling vier exemplaren van hun opstel produceerden. Docenten kregen formulieren waarmee zij snel aan leerlingen feedback konden geven over hun schrijfmappen. Ook werden per schrijfronde aangepaste formulieren geleverd, waarop leerlingen het gekregen commentaar konden ordenen tot een overzicht. Het complete lesmateriaal (De Schrijfmap) is te raadplegen in het bijlagenrapport 1, behorend bij dit onderzoeksverslag. In hoofdstuk 6 doen we nauwkeurig verslag van de gang van zaken van de schrijfronden, waardoor de lezer een beeld krijgt van onze aanpak.

Tabel 4.4.: Overzicht van lesmateriaal t.b.v. het experimentele stelonderwijsprogramma

| | 1e ronde | 2e ronde | 3e ronde | 4e ronde |
|---|---|---|---|---|
| Algemene informatie | 4 pagina's. | 5 pagina's. | 2 pagina's. | |
| Informatie over schrijven en opstellen | | | | |
| deel A | 23 pagina's over doel- en publiekgerichtheid. | 20 pagina's over structuur en betrouwbaarheid. | 16 pagina's over nieuwswaarde en taalgebruik. | 6 pagina's, bevattende een samenvattende schrijfprocedure. |
| deel B | 6 pagina's over de nadenkoefeningen. | 4 pagina's over de nadenkoefeningen. | | |
| deel C | 23 pagina's over 't waarom van commentaar geven, de problemen en de formulieren voor doelgerichtheid en publiekgerichtheid. 12 pagina's over het verwerken van commentaar. | 5 pagina's: de formulieren structuur en betrouwbaarheid. 5 pagina's over herschrijfplannen. | 4 pagina's: de formulieren nieuwswaarde en taalgebruik. | |
| Werkafdeling | 22 pagina's: de schrijfopdracht en een voorgedrukte complete schrijfronde. | 19 pagina's met informatie over het onderwerp. | 25 pagina's met informatie over het onderwerp. | 30 pagina's met informatie over het onderwerp. |

## 5. INSTRUMENTATIE EN MEETKWALITEIT VAN DE DATAVERZAMELING

In dit onderzoek onderscheiden we drie soorten afhankelijke variabelen: Schrijfprestatievariabelen, Schrijfprocesvariabelen en Psychologische Variabelen. Voor alle variabelen zijn meetinstrumenten ontwikkeld; sommige daarvan zijn in vooronderzoek beproefd. Voor de Schrijfprestatievariabelen zijn opstelschalen geconstrueerd (par. 5.1.) en analytische scoringsvoorschriften (par. 5.2.). Voor de Schrijfprocesvariabelen is een codeerschema ontwikkeld (par. 5.3.). Voor de Psychologische Variabelen is een zelfrapportage-instrument samengesteld (par. 5.4.). Van elk instrument wordt de constructie beschreven met beschikbare betrouwbaarheidsgegevens. Tevens worden meetkwaliteiten van de dataverzameling gerapporteerd. Voor de wijze van dataverzamelen, aantallen proefpersonen e.d. verwijzen we naar hoofdstuk 6.

### 5.1. Opstelschalen

Voor het maken van Schrijfprestatievariabelen werden twee schrijfopdrachten opgesteld, opstelschalen geconstrueerd en analytische scoringsvoorschriften ontwikkeld. In deze paragraaf wordt de constructie van de schrijfopdrachten en de opstelschalen beschreven, alsmede de kwaliteit van de via schaalbeoordeling verkregen scores uit het hoofdonderzoek.

Constructie. Voor het eerste en tweede toetsmoment werd één schrijfopdracht ontwikkeld, voor het derde moment een tweede. De schrijfopdrachten sloten qua opzet aan bij de schrijfopdrachten uit de leergang. De opdrachten bevatten: een situatieschets, bevattende de aanleiding, het doel, het publiek en het medium; een opdracht, met daarin nadrukkelijk nog eens vermeld het doel, het publiek en het medium; een hoeveelheid documentatie over het onderwerp, bestaande uit citaten en bewerkingen van fragmenten uit dagbladen, tijdschriften en boeken. De eerste opdracht werd ontleend aan een zogenaamde H.A.V.O.-A eindexamenopdracht. Leerlingen dienden een antwoord te formuleren en te beargumenteren over de vraag 'Bestaat er stress onder scholieren?'. Deze opdracht was op de proefschool alsmede in de finale beproeving (hoofdstuk 4) uitgeprobeerd. De tweede opdracht werd op eigen gezag ontwikkeld. Leerlingen moesten nu hun mening geven over 'De zin of onzin van huiswerk'. Deze opdracht werd beproefd op een zevental scholen in het kader van de constructie van een aantal meetinstrumenten voor taalvaardigheid voor een ander project (Van den Bergh et al., 1986).

Voor de constructie van de opstelschalen werd de door Blok en Hoeksma (1984) beproefde werkwijze gevolgd. Er werd een verzameling in kwaliteit verschillende en uiteenlopende opstellen uitgetypt en na een training aan een vijftal Neerlandici voorgelegd. Deze laatsten dienden te schatten hoeveel keer beter een te beoordelen opstel was dan een door de on-

derzoeker gekozen standaardopstel, dat door projectteamleden als middelmatig was beoordeeld. Deze taak werd voor elk van de vier prestatievariabelen verricht, aan de hand van globale kenmerkenlijstjes.

Per variabele werden twee tot vijf beoordelingsvragen gesteld, die werden vergezeld door opstelkenmerken die men in het te beoordelen opstel zou moeten aantreffen bij een positief respectievelijk negatief antwoord. De volgende beoordelingsvragen werden gesteld:

Publiekgerichtheid: De relatie tussen schrijver en lezer: maakt de schrijver contact met de lezer? De relatie tussen onderwerp en lezer: bevat de tekst inhoudelijke elementen waaruit blijkt dat de lezer met het onderwerp te maken heeft?

Doelgerichtheid: Bevat de tekst een duidelijk standpunt in de discussie? Bevat de tekst inhoudelijke elementen die de overtuigingskracht van de onderbouwing bevorderen/verminderen?

Opbouw: Is de tekst optisch goed ingedeeld? Is het opstel goed ingedeeld in inleiding, kern en slot? Wordt de hoofdgedachte in de inleiding en het slot geformuleerd? Zijn de alinea's goed met elkaar verbonden? Zijn de alinea's goed opgebouwd?

Stijl: Zijn de zinsbouw en de woordkeus gevarieerd? Bevat het opstel taaltrucjes, stijlfiguren? Is het taalgebruik persoonlijk?

Er werden 64 opstellen in getypte vorm in boekjes gebundeld; de volgorde in elk der boekjes was een willekeurige. Per beoordelaar waren er acht boekjes: voor elk van de vier variabelen twee: een boekje waarin de oordelen moesten worden gegeven in getallen en een boekje waarin de oordelen moesten worden gegeven in lijnen. De vijf beoordelaars drukten eerst in lijnen en later in getallen uit hoeveel beter of slechter dan het gegeven standaardopstel het te beoordelen opstel was (vgl. Blok en Hoeksma, 1984). De beoordelingszittingen vonden plaats op het onderzoeksinstituut. In het algemeen bleken de beoordelaars in deze eerste constructieperiode niet allen even goed gestuurd te worden door de instructies. De stabiliteit van de beoordelaars was over het algemeen redelijk, maar hun onderlinge overeenstemming liet te wensen over. Deze data bleken in eerste instantie geen solide basis voor een opstelschaal. Immers, van een in de schaal opgenomen opstel zou met veel te weinig zekerheid gezegd kunnen worden dat het een zeker schaalpunt representeerde. Alleen de gegevens voor de variabelen Stijl en Opbouw ('Huiswerk') werden voldoende geacht. Voor de overige zes schalen moesten opnieuw constructiezittingen worden gehouden.

Op drie wijzen werd getracht in de tweede constructieperiode de beoordelaars beter op elkaar af te stemmen: via beoordelaarsselectie, intensievere trainingen en feedback. Eén beoordelaar werd op grond van zijn afwijkend beoordelaarsgedrag niet voor een tweede constructieperiode uitgenodigd. De constructeurs ontvingen naast de beoordelingsvragen en kenmerkenlijstjes opstellen die als voorbeelden bij de kenmer-

kenlijstjes dienden. Deze werden in instructiebijeenkomsten besproken evenals het standaardopstel. Een tiental opstellen werd als training beoordeeld en besproken. Daarna vervulden de constructeurs onafhankelijk van elkaar hun beoordelingstaak. De constructeurs zeiden behoefte te hebben aan feedback. Regelmatig werd daarom de beoordeling onderbroken voor het berekenen van correlaties. Als een enkel opstel een lage correlatie veroorzaakte werd het besproken. De resultaten van deze tweede constructieperiode bleken wel voldoende als basis voor een schaal. In tabel 5.1. staan de resultaten van de tweede constructieperiode samengevat, alsmede de resultaten van de eerste periode ('Huiswerk', Stijl en Opbouw).

Tabel 5.1.: Stabiliteitscoëfficiënten[1] (s) en compatibiliteitscoëfficiënten[2] (c) bij de schaalconstructie voor de opdrachten 'Huiswerk' (H) en 'Stress' (S) na logaritmische transformatie der ruwe scores

| Beoordelaar | Publiekgerichtheid* | | | | Doelgerichtheid** | | | | Opbouw* | | | | Stijl* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | s | | c | | s | | c | | s | | c | | s | | c | |
| | H | S | H | S | H | S | H | S | H | S | H | S | H | S | H | S |
| 1 | .76 | .63 | .71 | .73 | .72 | .83 | .81 | .76 | .78 | .77 | .81 | .82 | .79 | .70 | .71 | .70 |
| 2 | .78 | .83 | .74 | .74 | .83 | .94 | .78 | .82 | .86 | .89 | .77 | .87 | .81 | .86 | .79 | .79 |
| 3 | .88 | .81 | .81 | .76 | .96 | .93 | .71 | .88 | .87 | .88 | .87 | .88 | .85 | .78 | .88 | .73 |
| 4 | .89 | .74 | .84 | .83 | .85 | .69 | .82 | .77 | .84 | .81 | .72 | .86 | .78 | .68 | .73 | .62 |

1) Stabiliteit: de correlatie tussen de scores van een beoordelaar op 'lijnen' en op 'getallen'.
2) Compatabiliteit: de correlatie tussen de scores van één beoordelaar en de overige beoordelaars.
* n = 64
** n = 30

Vervolgens stelden we schalen samen waarmee de opstellen uit het experiment beoordeeld werden. De scores van 'lijnen' en 'getallen' werden logaritmisch getransformeerd om een intervalschaal te verkrijgen en vervolgens per beoordelaar gesommeerd. Deze werden op hun beurt gesommeerd tot juryscores. De homogeniteit van deze scores bleek hoog (zie tabel 5.2.).

Tabel 5.2.: Homogeniteit van jury-oordelen bij opstelschaalconstructie (Cronbachs alpha; n=4 beoordelaars)

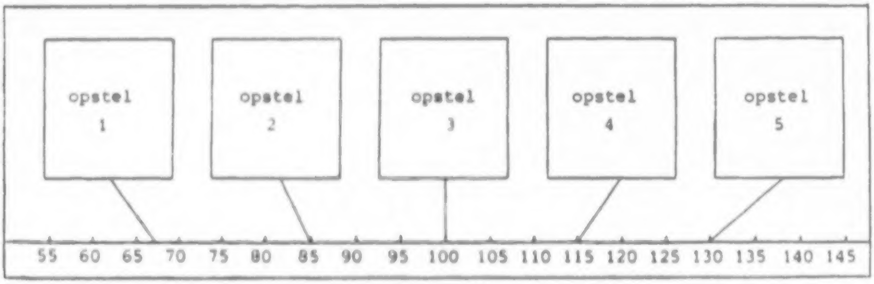
| Opdracht | Publiekgerichtheid | Doelgerichtheid | Opbouw | Stijl |
|---|---|---|---|---|
| Huiswerk | .96 | .95 | .90 | .90 |
| Stress | .90 | .92 | .94 | .86 |

Deze juryscores werden getransformeerd naar een schaal met een gemiddelde van 100 en een standaardafwijking van 15. Van elk van de acht schalen is de gemiddelde standaardafwijking van de beoordelingen berekend; deze beliep waarden van 2.8 tot 4.4. Deze waarden zijn een indicatie voor de zekerheid waarmee men weet dat de score een zekere waarde op de schaal representeert.

Tabel 5.3.: Correlaties tussen de opstelschalen voor Publiekgerichtheid, Doelgerichtheid, Opbouw en Stijl. Boven de diagonaal voor de opdracht 'Huiswerk', onder de diagonaal voor de opdracht 'Stress'

| Schaal | Publiekgerichtheid | Doelgerichtheid | Opbouw | Stijl |
|---|---|---|---|---|
| Publiekgerichtheid | - | .39 | .17 | .66 |
| Doelgerichtheid | .17 | - | .17 | .10 |
| Opbouw | .54 | .03 | - | .44 |
| Stijl | .62 | .31 | .45 | - |

Uit de acht opstelschalen werden voor elke schaal vijf opstellen gekozen die op of nabij de schaalpunten 70, 85, 100, 115 en 130 lagen. Alleen opstellen waarover beoordelaars het relatief sterk met elkaar eens waren, kwamen voor selectie in aanmerking. Het resultaat van deze keuze moest bovendien aan de voorwaarde voldoen dat elke schaal andere opstellen bevatte. We beschikten tenslotte over acht schalen, met ieder vijf voorbeeldopstellen, waarmee de beoordelingstaak verricht werd. Deze beoordelingsschalen zijn integraal opgenomen in het tweede bijlagenrapport. Een beoordelingsschaal bestond uit een groot vel papier, waarop een lijn aangebracht was met als grenspunten 70 en 130. Op gelijke afstanden waren 85, 100 en 115 op de lijn aangebracht. Boven de lijn waren de vijf opstellen ingeplakt die elk een schaalpunt representeerden. In schema:



Dataverzameling. Uit het hoofdonderzoek resulteerden 792 opstellen (zie hoofdstuk 6). Deze werden door twee onafhankelijke beoordelaars op de variabelen Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl beoordeeld met behulp van opstelschalen. Beoordelaars dienden aan de te beoordelen opstellen scores toe te kennen na vergelijking van de kwaliteit van het te beoordelen opstel en de voorbeelden van de opstelschalen. Twee beoordelaars van wie bekend was uit ander aan de S.C.O. verricht onderzoek dat zij nauwkeurig beoordeelden en onderling overeenstemden, verrichtten de beoordelingstaak voor achtereenvolgens de aspecten Opbouw, Doelgerichtheid en Stijl. Zij verrichtten de taak op het instituut. Twee andere beoordelaars, die ook bij de constructie van de opstelschalen betrokken waren, beoordeelden de opstellen op de variabele Publiekgerichtheid. Eén van beiden beoordeelde de opstellen thuis. Voor elke variabele werden eerst de opstellen 'Stress' (voor- en tussentoetsopstellen waren onherkenbaar voor beoordelaars gemixed) en vervolgens de opstellen 'Huiswerk' (natoets) beoordeeld. De opstellen waren in willekeurige volgorde gelegd, maar wel zo dat de opstellen van een klas zich op gelijke afstand bevonden, opdat geen klas het voor- of nadeel had zich voor- of achterin de reeks van 792 opstellen te bevinden. Het hele pakket werd in porties van 15 opstellen verdeeld. Elke beoordelaar beoordeelde de porties en dus de opstellen in dezelfde volgorde. Voorafgaand aan de beoordelingszittingen werd een proefset thuis als voorbereiding beoordeeld en werden de voorbeeldopstellen op de schaal geanalyseerd. Daarop volgde een bespreking onder leiding van de onderzoeker. Er werd gestreefd naar consensus over de waarde van de schaalvoorbeelden en over de inhoud van de beoordelingsvragen. Tijdens beoordelingszittingen werden steekproeven genomen om de overeenstemming tussen de beoordelaars vast te stellen. Als dat naar aanleiding van een relatief lage correlatie ( .50) nodig bleek, werden opstellen die die lage correlatie veroorzaakten, opgespoord. Een paar daarvan werden door de beoordelaars besproken onder leiding van de projectleider. Dit leidde tot steeds verfijndere beslissingsstrategieën voor de beoordelaars. Na afloop van het beoordelen van een variabele herbeoordeelden de beoordelaars nog een dertigtal opstellen opdat hun stabiliteit berekend kon worden. De betrouwbaarheidsgegevens zijn vermeld in tabel 5.4.

De variabelen bleken vrij onafhankelijk. De correlaties tussen de variabelen waren niet hoog (tabel 5.5.).

Tabel 5.4.: Stabiliteitscoëfficiënten (Pearson correlaties tussen eerste en tweede beoordeling) van elk der beide beoordelaars en de overeenstemmingscoëfficiënten (Pearson correlaties) tussen beide beoordelaars van hun beoordelingen met behulp van opstelschalen

| Variabele | Stabiliteit (n = 30) | | Overeenstemming (n = 792) |
|---|---|---|---|
| Publiekgerichtheid | .81 | .84 | .64 |
| Doelgerichtheid | .89 | .91 | .78 |
| Opbouw | .80 | .69 | .78 |
| Stijl | .89 | .97 | .76 |

Tabel 5.5.: Correlaties tussen de beoordelingen van de Schrijfprestatievariabelen Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl op drie toetsmomenten (n = 264)

| | | Opbouw | Doelger. | Publiekger. |
|---|---|---|---|---|
| | Doelgerichtheid | .47 | | |
| Tijdstip 1 | Publiekgerichtheid | -.03 | .13 | |
| (n=264) | Stijl | .38 | .29 | .27 |
| | | Opbouw | Doelger. | Publiekger. |
| | Doelgerichtheid | .34 | | |
| Tijdstip 2 | Publiekgerichtheid | .30 | .28 | |
| (n=264) | Stijl | .52 | .40 | .54 |
| | | Opbouw | Doelger. | Publiekger. |
| | Doelgerichtheid | .58 | | |
| Tijdstip 3 | Publiekgerichtheid | .34 | .31 | |
| (n=264) | Stijl | .43 | .30 | .44 |

## 5.2. Analytische opstelscoringsvoorschriften

Twee Schrijfprestatievariabelen, Doel- en Publiekgerichtheid, werden op nog andere wijze dan via opstelschalen gemeten. Voor deze variabelen werden analytische scoringsvoorschriften ontwikkeld. Het leek verstandig om deze twee belangrijke variabelen op twee wijzen te operationaliseren. Operationalisaties op analytisch niveau waren nog niet beschikbaar. Een poging daartoe zou ook van didactisch belang kunnen zijn.

Constructie. Er werd een beperkte literatuurstudie verricht (Jansen en Scheringa, 1984). De daaruit verkregen eisen die aan een doel- en publiekgerichte tekst zouden moeten worden gesteld, werden gecategoriseerd. Deze categorieën dienden als hulpmiddel voor het genereren van scoringsitems. Gestreefd werd naar een dichotome scoring. De scoringsvoorschriften zijn door voor de scoringstaak geworven Neerlandici beproefd. Op

basis hiervan zijn de formuleringen bijgesteld. Tevens werden aan de items toelichtingen toegevoegd, alsmede een aantal opstellen waarnaar verwezen werd in genoemde toelichtingen.

Het object van de scoringen van beide opstelkwaliteiten was verschillend. Voor de scoring van Doelgerichtheid was de hele tekst uitgangspunt. Voor Publiekgerichtheid werd de inleiding van het opstel gescoord. Scoring van de hele tekst zou een te omvangrijke taak zijn. Bovendien is de inleiding bij uitstek het tekstgedeelte, dat vrijwel uitsluitend tot doel heeft een lezer te bereiken (zie voor de relationele functie Braet, 1979; Schulz von Thun et al., 1982). Dit relationele aspect van Publiekgerichtheid werd in de experimentele leergang sterk benadrukt.

Het bleek mogelijk voor Publiekgerichtheid scoringsitems te formuleren, die onafhankelijk van elkaar gescoord konden worden, d.w.z. in principe in elke willekeurige volgorde (op item 19 na). Voor Doelgerichtheid bleek dit niet mogelijk: sommige beslissingen waren afhankelijk van elkaar. In tabel 5.6. worden de scoringsitems voor Doelgerichtheid gepresenteerd, in tabel 5.7. die voor Publiekgerichtheid.

Tabel 5.6.: De items van de scoringsvoorschriften Doelgerichtheid. Voor het item is het itemnummer vermeld

| Categorieën | Items |
|---|---|
| Tekstsoort | 1. Is de tekst overwegend argumentatief? |
| Standpunt | 2. Formuleert de schrijver een standpunt over 'Bestaat er stress onder scholieren?' / 'Heeft huiswerk zin?'<br>3. Op welke plaats in de tekst formuleert de schrijver expliciet zijn standpunt (inleiding, kern, slot)? |
| Onderbouwing vam het standpunt | 4. Wordt een poging gedaan het standpunt aannemelijk te maken?<br>5. Hoe onderbouwt de schrijver zijn standpunt?<br>6. Is de onderbouwing voldoende om het standpunt aanvaardbaar te maken? |
| Samenhang | 7. Is het betoog over het bestaan van het verschijnsel stress onder scholieren / de zin van huiswerk het centrale element van de tekst?<br>8. Is de informatie van de tekst zodanig gerangschikt dat er één heldere lijn in het opstel zit? |

Tabel 5.7.: De items van de scoringsvoorschriften Publiekgerichtheid van de inleiding. Voor het item is het itemnummer vermeld

---

HOOFDCATEGORIE INHOUD

Subcategorie Contact (Schrijver-Lezer)

1. Wordt de lezer persoonlijk aangesproken (je, jij, jullie, gebiedende wijs)?
2. Manifesteert de schrijver zichzelf (ik, me)?
3. Kenmerkt de schrijver zichzelf?
4. Suggereert de schrijver overeenkomsten tussen zichzelf en lezers?

Subcategorie Appèl (Lezer-Onderwerp)

5. Wijst de schrijver via een expliciete feitelijke uitspraak op het belang van de informatie voor de lezer?
6. Roept de schrijver emoties op bij de lezer?

Subcategorie Aandacht/Spanning

7. Stelt de schrijver een vraag?
8. Presenteert de schrijver het onderwerp als iets ongewoons?
9. Gebruikt de schrijver beeldspraak die aansluit bij de interesse/ervaring van de lezer?
10. Roept de schrijver iets uit?
11. Vertelt de schrijver een anekdote?
12. Wordt de aandacht van de lezer op een andere (inhoudelijke) manier getrokken dan 7-11?

HOOFDCATEGORIE BEGRIJPELIJKHEID

Subcategorie Structuur

13. Bevat de tekst een titel die inhoudelijk en/of structureel de lezer voorbereidt?
14. Beschrijft de schrijver het communicatieve kader waarin de tekst te plaatsen is?
15. Identificeert de schrijver het tekstthema?
16. Verstrekt de schrijver informatie over de rest van de tekst?
17. Bevat de tekst uitsluitend elementen die relevant zijn voor een publiekgerichte inleiding?

Subcategorie Taalgebruik

18. Bevat het fragment één of meer zinnen langer dan 15 woorden?
19. Zijn deze zinnen allemaal goed leesbaar?
20. Bevat de tekst een moeilijk abstract woord?

---

Dataverzameling. De scoring van de 792 opstellen werd verricht door vijf werkloze Neerlandici, die ook bij de opstelschaalconstructie betrokken waren. Ieder scoorde een pakket van 200 opstellen. Dit bestond uit 148 voor elke scoorder verschillende opstellen en 52 gemeenschappelijke opstellen. De verdeling van de opstellen was zodanig dat elke scoorder nagenoeg evenzeer bijdroeg aan de te construeren klassescore. De volgorde waarin de 148 opstellen werden aangeboden was een willekeurige. De set van 52 opstellen was op een voor elke scoorder gelijke volgorde regelmatig over het pakket van 148 verspreid. Deze 52 opstellen werden na afloop van de scoring nogmaals gescoord door elke beoordelaar. Nadat de scoring en herscoring voor de variabele Doelgerichtheid had plaatsgevonden, werd de scoring en herscoring voor Publiekgerichtheid afgewerkt. De beoordelaars werkten onafhankelijk van elkaar.

De gemeenschappelijke set van 52 opstellen was de basis voor betrouwbaarheidsanalyses. Daartoe werden eerst enkele hercoderingen aangebracht, zodat alle scoringsitems dichotoom werden. Een aantal scoringsitems van Doelgerichtheid was afhankelijk. Dit maakte het onmogelijk om de gebruikelijke toets- en itemanalyses toe te passen, waarmee de homogeniteit van de toets kon worden bepaald. Voor de variabele Publiekgerichtheid kon dit wel (KR-20: .24). Beide verzamelingen items konden derhalve niet als een homogene test beschouwd worden. Besloten werd geen somscores te berekenen en uitsluitend op itemniveau, mits het betreffende item betrouwbaar gescoord was, exploratief naar veranderingen tussen toetsmomenten en condities te kijken.

Per scoringsitem werd de stabiliteit van elke beoordelaar berekend en de overeenstemming tussen beoordelaars. Voor de overeenstemming is Cronbachs alpha berekend. Op dichotome items toegepast komt deze index overeen met KR-20. Voor de stabiliteit is een index berekend, bekend als Bennetts S (Bennett, Blomquist en Goldstein, 1954). In deze index wordt de ruwe overeenstemming gecorrigeerd voor de kans die een observatie heeft om door toeval in een categorie terecht te komen. Brennan en Prediger (1981) spreken in hun overzicht van overeenstemmingsmaten de voorkeur uit voor Bennetts S als de marginalen niet van tevoren bepaald zijn. In de tabellen 5.8. en 5.9. worden de betrouwbaarheidsresultaten van de scoringen voor respectievelijk Doelgerichtheid en Publiekgerichtheid gepresenteerd. Uit de tabellen blijkt dat we er maar zeer gedeeltelijk in geslaagd zijn voor Doelgerichtheid betrouwbaar te scoren items te formuleren; voor Publiekgerichtheid is dat beter gelukt. Bij de interpretatie van de resultaten (par. 7.5.) zullen we aan de scores op onbetrouwbaar gescoorde items weinig waarde kunnen hechten. Voor Doelgerichtheid gaat het dan om de items 4,5a,5b,5c,5d,5e en 8; voor Publiekgerichtheid om item 6.

Tabel 5.8.: Stabiliteiten (Bennetts S) van vijf beoordelaars en de jurybetrouwbaarheid (Cronbachs alpha) van de analytische scoring van de variabele Doelgerichtheid. (n varieert, omdat scoringsitems afhankelijk van elkaar zijn en is maximaal 52)

| Scoringsitem | Stabiliteiten | | | | | Jurybetrouwbaarheid |
|---|---|---|---|---|---|---|
| 1 | .33 | .57 | .54 | .54 | .56 | .68 |
| 2 | .76 | .75 | .96 | .88 | .81 | .86 |
| 3a | .61 | .53 | .66 | .83 | .82 | .94 |
| 3b | .61 | .69 | .51 | .83 | .76 | .84 |
| 3c | .81 | .76 | .61 | .92 | .70 | .86 |
| 4 | .66 | .88 | .90 | .75 | a | .13*** |
| 5a | .90 | a | .95 | .58 | .94 | .22** |
| 5b | .62 | .47 | .27 | .79 | .56 | .72 |
| 5c | .62 | .83 | .37 | .58 | .56 | -.03 |
| 5d | .52 | .13 | .76 | .79 | .69 | .32* |
| 5e | .90 | .80 | .66 | .79 | .69 | -.51 |
| 5f | .52 | .100 | .51 | .26 | .88 | .75* |
| 6 | .38 | .60 | .71 | .75 | .06 | .58 |
| 7 | .37 | .58 | .54 | .62 | .45 | .57 |
| 8 | .41 | .38 | .88 | .77 | .29 | .33 |

* = de scores van één beoordelaar vertoonden geen variantie.
** = de scores van twee beoordelaars vertoonden geen variantie.
*** = de scores van drie beoordelaars vertoonden geen variantie.
a) Voor deze beoordelaars kon op dit item geen Bennetts S berekend worden wegens een gebrek aan variantie doordat zowel op tijdstip 1 als op tijdstip 2 alle opstellen eenzelfde code kregen.

Tabel 5.9.: Stabiliteiten (Bennetts S) van vijf beoordelaars en jurybetrouwbaarheid (Cronbachs alpha) van de scoring van de variabele Publiekgerichtheid. (n=52, m.u.v. item 12 en 19 die afhankelijk zijn; n=36)

| Scoringsitem | Stabiliteiten | | | | | Jurybetrouwbaarheid |
|---|---|---|---|---|---|---|
| 1a | .84 | .88 | .92 | .100 | .80 | .92** |
| 1b | .100 | a | .96 | .100 | a | .77 |
| 1c | .69 | .88 | .88 | .92 | .84 | .86 |
| 1d | .80 | .100 | .85 | .96 | .88 | .97 |
| 2 | .92 | .96 | .100 | .100 | .80 | .99 |
| 3 | .100 | .92 | .92 | .92 | .92 | .82 |
| 4 | .92 | .88 | .92 | .92 | .84 | .88 |
| 5 | .65 | .77 | .73 | .77 | .88 | .67*** |
| 6 | a | a | .77 | a | .100 | .25 |
| 7 | .92 | .96 | .92 | .100 | .76 | .97* |
| 8 | .96 | .100 | .96 | .96 | .100 | .90** |
| 9 | a | .96 | .100 | .100 | a | .62 |
| 10 | .96 | .100 | .100 | .100 | .92 | .92* |
| 11 | .84 | .88 | .73 | .96 | .96 | .47 |
| 12 | .76 | .88 | .79 | .77 | .96 | .41 |
| 13 | .88 | .100 | .92 | .96 | .92 | .97** |
| 14 | .96 | .100 | .96 | a | .100 | .83 |
| 15 | .53 | .62 | .25 | .81 | .61 | .67 |
| 16 | .84 | .92 | .85 | .88 | .88 | .77 |
| 17 | .88 | .69 | .58 | .65 | .53 | .71 |
| 18 | .88 | .96 | .72 | .100 | .96 | .94 |
| 19 | .65 | .63 | .49 | .77 | .53 | .70* |
| 20a | .100 | .88 | .92 | .92 | a | .72 |
| 20b | .96 | .88 | .85 | .88 | .100 | .73* |
| 20c | .96 | .85 | .85 | .85 | a | .53 |

* = de scores van één beoordelaar vertoonden geen variantie.
** = de scores van twee beoordelaars vertoonden geen variantie.
*** = de scores van drie beoordelaars vertoonden geen variantie.
a) Voor deze beoordelaars kon op dit item geen Bennetts S berekend worden wegens een gebrek aan variantie doordat zowel op tijdstip 1 als op tijdstip 2 alle opstellen eenzelfde code kregen.

## 5.3. Schrijfprocesvariabelen

Voor het beschouwen van verschillen in schrijfprocessen is gekozen voor de methode van hardop denken als dataverzamelingstechniek (Ericsson en Simon, 1984). Er werd op basis van het schrijfprocesmodel van Hayes en Flower (1980, 1983) een codeerschema ontwikkeld, aangepast aan de hypothesen die wij

in hoofdstuk 2 formuleerden. In deze paragraaf beschrijven we het codeerschema en de kwaliteit van de met behulp van dit schema verkregen gegevens in termen van betrouwbaarheid. Het schema is integraal opgenomen in bijlagenrapport 2.

Constructie. Zoals in hoofdstuk 2 beschreven is, onderscheiden Hayes en Flower (1980; 1983) drie hoofdprocessen: het plannings-, het formuleer- en het reviseerproces. De afstemming van de processen wordt geregeld door een monitor. Twee van genoemde processen bestaan weer uit subprocessen. Planning bestaat uit genereren, structureren en doelbepalen. Reviseren bestaat uit lezen en corrigeren. Deze hoofd- en onderverdeling vormde de basis voor ons codeerschema. Het codeerschema werd op grond van een aantal proefprotocollen, de specifieke schrijfopracht en onze hypothesen verfijnd. Het uiteindelijke doel was zoveel mogelijk verschillende (sub-)processen in het schema onder te brengen met het oog op latere analyse. In het kader van de exploraties in dit onderzoek, beperken we de beschrijving van het schema tot de categorieën waarmee de hypothesen onderzocht zouden worden.

Daartoe moesten in de categorie genereren het genereren van ideeën en het genereren van presentatiestructuren worden onderscheiden; in de categorie doelbepalen het herhalen van taakeisen uit de opdracht en het genereren van eigen criteria voor een goede tekst; in de categorie formuleren het omzetten van gedachten in schrijftaal, - het zichzelf dicteren -, en het in overweging nemen van woorden en zinnen; en in de categorie corrigeren maakten we onderscheid in het veranderen van kleine tekstdelen (woorden, zinnen) en van grotere fragmenten. Om te kunnen bepalen of schrijvers gebruik maakten van interactionele monitoring (Hagen, 1981), brachten we in de categorieën genereren, doelbepalen en corrigeren het onderscheid aan tussen neutrale opmerkingen en opmerkingen waaruit een lezersperspectief bleek. In tabel 5.10. wordt het relevante gedeelte uit het codeerschema gepresenteerd. Het schema is via het coderen van proefprotocollen uitgeprobeerd, bijgesteld en uitgebreid (Baltzer en Bochardt, 1985).

Dataverzameling. Voor de wijze waarop de protocollen verzameld zijn, verwijzen we naar hoofdstuk 6. Nu gaan we alleen in op de wijze waarop de 44 protocollen (22 klassen, één leerling per klas, voor- en nameting) bewerkt werden en gescoord. De bandjes werden uitgetypt met behulp van een tekstverwerker, waarin ingevoerd waren de schrijfopdracht, de informatie over het onderwerp, de kladversie van het opstel en de netversie. Alles wat een leerling volgens het bandje ertussendoor zei, werd aan het protocol toegevoegd en wat niet hardop werd gezegd van wat al in de tekstverwerker stond, werd gewist. De typist bracht in de marges al vier codes aan: hardop schrijven, hardop lezen van de opdracht, hardop lezen van eigen tekst en hardop denken. Stiltes werden met puntjes aangegeven. Er werd geen interpunctie aangebracht: als een proefpersoon een gedachte afrondde volgens het intonatiepatroon, werd een schuine streep getypt.

Tabel 5.10.: Codeerschema schrijfprocessen (fragment uit Baltzer en Bochardt, 1985)

| Processen | Voorbeelden van protocolfragmenten |
|---|---|
| 02. Zelfinstructie t.a.v. proces/aanpak | |
| 1. voor het schrijven van tekst (ook overschrijven) | ik ga nu eerst maar eens de voorbeelden opschrijven |
| 2. voor het lezen van eigen tekst | even overlezen wat ik nu heb staan |
| 3. voor het lezen van opdracht/citaten | ik ga de citaten nog maar eens lezen |
| 4. voor het genereren (bedenken) van inhoud of vorm | ik ga eerst bedenken wat ik er allemaal in zou kunnen schrijven/even een titel bedenken / even een slot bedenken |
| 5. voor structureren van ideeën, ervaringen, etc. | wat ik nu heb moet ik gaan ordenen |
| 6. voor reviseren van tekst | ik ga het doorlezen op spelfouten |
| 7. voor de aanpak van de schrijfopdracht als geheel of 'n aanzienlijk deel van de schrijfopdracht | ik maak eerst een schema op klad en dan begin ik meteen in het net te schrijven / ik ga nu in het net schrijven |
| 8. voor het stoppen met schrijven/lezen | ik stop er nu mee / dat was het / klaar! |
| | dat doe ik nu nier verder / dat doe ik straks wel |
| 03. Het stellen of opnieuw stellen van produktdoelen of criteria | |
| 1. taakeisen herhalen (opnieuw lezen, inprenten) — N (=neutraal) | het moeten dus ongeveer 2 kantjes worden |
| — P (=publiekgericht) | en ik moet medeleerlingen overtuigen |
| 2. eigen doelen of criteria weergeven waaraan de tekst moet voldoen — N = 1 | het moeten niet te lange zinnen worden |
| — P = 2 | er moet een voorbeeld in dat leerlingen leuk vinden |
| 04. Genereren | |
| 1. ideeën, ervaringen, kennis, meningen (inhouden voor het opstel) — N = 1 | een nadeel is misschien ook dat het leven met kinderen veel duurder is |
| — P = 2 | mondelinge beurten, dat is een voorbeeld van stress dat ze wel zullen herkennen |
| (ook opmerkingen die erop wijzen dat men met genereren bezig is, zonder al specifieke inhouden te noemen) | wat zal ik daar over zeggen / waar zal ik mee beginnen |
| 2. alternatieve presentatiestructuren (het betreft hier afwegen en overdenken, niet het stellen als norm of doel) — N = 1 | ik kan er natuurlijk een verhaaltje van maken |
| — P = 2 | met voorbeelden uit de praktijk is het leuker om te lezen |
| (ook niet-inhoudelijke opmerkingen) | hoe zal ik dat eens opschrijven |
| 08. al schrijvende keuzes maken voor | |
| 1. spelling | stress...met dubbel s...bij...scholieren |
| 2. zinsbouw | hoe...zou dat komen....komt dit... |
| 3. woorden | leerlingen...zouden...zich...dienstbaar, nee nuttig... kunnen maken..... |
| 4. interpunctie | komma? nee, dubbele punt |
| 5. layout | nieuwe alinea / nieuwe regel / regel overslaan |
| 10. evalueren (heroverwegen/beoordelen) van | (hier geldt hetzelfde als bij 'lezen') |
| 1. één of een aantal woorden die het laatst zijn opgeschreven — N = 1 / P = 2 | neutrale opmerkingen zijn bijv.: dit slaat nergens op |
| | dit heeft niets met stress te maken |
| | dit is een rare zin |
| 2. één of een aantal woorden verder terug in de tekst — N = 1 / P = 2 | publiekgerichte opmerkingen zijn bijv.: |
| 3. het tekstgedeelte dat het laatst is opgeschreven — N = 1 / P = 2 | dit begrijpen ze niet |
| | dit zullen ze wel saai vinden |
| 4. een tekstgedeelte verder terug in de tekst — N = 1 / P = 2 | |
| 5. de hele tot dan toe geproduceerde tekst — N = 1 / P = 2 | |
| 6. de interpunctie | |
| 7. heroverwegen wat in het klad staat, wanneer men bezig is in het net te schrijven | |

vervolg tabel 5.10: Codeerschema schrijfprocessen (fragment)

11. <u>veranderen</u> (verbeteren) of schrappen of toevoegen

1. van één of een aantal woorden die het laatst zijn opgeschreven
   - N = 1
   - P = 2
2. van één of een aantal woorden verder terug in de tekst
   - N = 1
   - P = 2
3. het tekstgedeelte (passage) dat het laatst is opgeschreven
   - N = 1
   - P = 2
4. een tekstgedeelte (passage) verder terug in de tekst
   - N = 1
   - P = 2
5. de hele tot dan toe geproduceerde tekst
   - N = 1
   - P = 2
6. de interpunctie
7. in het 'net' verbeterd overschrijven van het 'klad', wat betreft (<u>i.p.v. N of P</u>):
   1. spelling
   2. zinsbouw
   3. woordkeus
   4. een (bij)zin of passage toevoegen
   5. een (bij)zin of passage weglaten
   6. een woord toevoegen
   7. een woord weglaten
   8. besluiten iets niet te verbeteren ('nou ja, laat maar zitten')

neutrale opmerkingen zijn bijv.:
dit streep ik door
van deze zin maak ik twee korte zinnen

een publieksgerichte opmerking is:
hier maak ik 'nuttig' van, dat snappen ze beter

---

De protocollen werden daarna ingedeeld in zinvolle eenheden (Wouters en De Jong, 1982). In een zinvolle eenheid vindt één cognitief proces plaats. De indeling is door twee assistenten uitgevoerd.

De protocolfragmenten zijn door twee Neerlandici gecodeerd. Beiden scoorden twee protocollen als training en bespraken de resultaten daarvan onder leiding van de onderzoeksassistent fragment voor fragment. Elk protocol is daarna door een van beiden gecodeerd. Het was onmogelijk voor hen te weten uit welke conditie het protocol afkomstig was. Gedurende de twee maanden dat er gecodeerd werd, is op vier momenten een protocol door beiden gecodeerd. Over deze protocollen is de betrouwbaarheid berekend met behulp van loglineaire modelpassing met protocol, codeurs en categorieën als termen. Het model zonder codeurs bleek goed te passen ($G^2$=20.42, df 32, p $\geq$.5). Daaruit kan geconcludeerd worden dat de scores verkregen met behulp van dit schema codeursonafhankelijk zijn.

Op grond van de datamatrices moesten we besluiten een aantal categorieën te laten vallen vanwege zeer lage celfrequenties. Het betrof het onderscheid tussen neutrale en publiekgerichte opmerkingen, het genereren van criteria en het genereren van presentatiestructuren. Deze scoringen werden opgenomen in een restcategorie. Hierdoor vervielen de hypothesen over de publiekgerichte aanpak van de proefpersonen in de experimentele conditie, over hun gebruik van criteria in het proces en over het genereren van presentatiestructuren (zie par. 2.5.)

## 5.4. Psychologische Variabelen

Uit de didactische literatuur, maar ook uit onderzoeken bleek dat men verwachtte dat een steldidactiek waarin leerlingen elkaars teksten becommentarieerden, vermindering van schrijfangst teweeg zou brengen. In ons onderzoek was het in eerste instantie ook de bedoeling de mate van schrijfangst als afhankelijke variabele op te nemen. Echter, na statistische analyse van een onder vijftienjarigen afgenomen vertaling van een Amerikaans schrijfangstinstrument (Writing Apprehension Test van Miller en Daly, 1975) bleek een verfijning van de afhankelijke variabelen mogelijk, die correspondeerde met in didactische literatuur (Rijlaarsdam, 1984) genoemde effecten. Met gebruikmaking van Bergen (1981) werd getracht een meer theoretisch onderbouwd instrument te ontwikkelen. Na een beproeving en statistische analyse werd het uiteindelijke instrument vastgesteld. In het navolgende beschrijven we de ontwikkeling gedetailleerder.

Constructie. Begonnen werd met de afname van een vertaling van Miller en Daly's W.A.T. De lijst van 26 items werd afgenomen op een aantal leerlingen die tot onze doelgroep behoorden (Atheneum en H.A.V.O., derde klassen, n=114). De scores werden geanalyseerd op inhoudelijke samenhang. De hele schaal bleek homogeen (Cronbachs alpha .86). Binnen de schaal bleken drie homogene, interpreteerbare schalen van elk vijf à zes items te onderscheiden, die niet al te sterk met elkaar correleerden (.42 - .66). Deze schalen benoemden wij naar de inhoud van de items die de schaal droegen. Onderscheiden werden de schaal Angst-voor-het-niet-Kunnen-schrijven (Cronbachs alpha .82, 5 items, kenmerkend item 'Ik heb, voordat ik aan de opstel begin, het idee dat het me slecht af zal gaan'), de schaal Houding-jegens-Schrijven (Cronbachs alpha .84, 6 items, kenmerkend item 'Schrijven is erg leuk') en de schaal Houding-jegens-Beoordeeld-worden (Cronbachs alpha .76, 6 items, kenmerkend item 'Ik zou het leuk vinden wat ik schrijf aan tijdschriften te sturen voor publikatie').

De gevonden subschalen stelden ons in staat specifiekere hypothesen te formuleren en te toetsen. Het leek ons wenselijk om binnen de beperkte mogelijkheden die het hier gerapporteerde onderzoek ons bood een poging te doen een meer theoretisch gefundeerd schrijfangstinstrument te ontwerpen. We zoch-

ten daarvoor aansluiting bij Bergen (1981), die de S.S.A.T. ontwierp, de Situatie Specifieke Angst Test. In navolging van een aantal theoretici onderscheidt Bergen drie componenten in angst: een cognitieve, een affectieve en een psychofysiologische component. Bergen beschrijft deze componenten als volgt.

Cognitieve component: dit is 'Angst-voor-het-niet-Kunnen' in de betekenis van verwachting, het zich zorgen maken of twijfelen aan een goede afloop van de te verrichten taak. Deze angst komt voort uit een laag concept van eigen bekwaamheid en is het resultaat van de subjectieve taxatie van de situatie die als bedreigend voor de persoon wordt ervaren. De items die Bergen formuleerde hebben betrekking op ervaringen van twijfel, onzekerheid, spanning en op ongerustheid over de eigen competentie.

Affectieve component: dit is wat Bergen met betrekking tot evaluatie-angst 'Evaluatieve gevoelens' noemt. Het gaat om ego-betrokken gevoelens, die de uitvoering van de taak negatief beïnvloeden. Situaties die positieve affecten van trots en tevredenheid kunnen opleveren worden opgezocht. Situaties die negatieve gevoelens als beschaamdheid, ergernis en teleurstelling behelzen, worden vermeden.

Psychofysiologische component: het betreft hier fysiologische (hartkloppingen, zweten, ademhalingsproblemen, buikpijn, beven) en psychische verschijnselen (in paniek raken, zenuwachtig zijn, niet kunnen inslapen) die optreden in situaties die voor de persoon bedreigend zijn. Ze maken hem bewust van gevoelens van spanning en van verhoogde activiteit van het autonoom zenuwsysteem. Deze ervaringen hinderen hem bij de effectieve uitvoering van taken.

Twee van de door ons aangetroffen clusters in de W.A.T. correspondeerden inhoudelijk met twee componenten die Bergen onderscheidde. Angst-voor-het-niet-Kunnen-schrijven was op te vatten als de cognitieve component van schrijfangst en Houding-jegens-Schrijven als de affectieve component. Besloten werd te pogen een schrijfangstinstrument te ontwerpen op basis van de drie door Bergen onderscheiden componenten. Daaraan werd toegevoegd een schaal voor het meten van de Houding-jegens-Beoordeeld worden. Bovendien werd een schaal opgenomen die we, analoog aan Bergens instrument, Persoonlijke Inzet noemden, om daarmee na te kunnen gaan of er een grotere motivatie om een goede tekst te schrijven gekweekt werd door een experimentele leergang zoals men in didactische literatuur wel aannam (Rijlaarsdam, 1984, p.25,29).

Bergens items voor de schalen voor de drie angstcomponenten en persoonlijke inzet, alsmede de items uit de W.A.T. dienden als voorbeeld voor de constructie van het nieuwe instrument. Dat bestond uit vijf schalen:
Cognitieve Component: Angst-voor-het-niet-Kunnen-schrijven (14 items AK);
Psychofysiologische Component: Begeleidingsverschijnselen (11 items BV);

Affectieve Component: Houding-jegens-Schrijven (10 items HS); Houding-jegens-Beoordeeld-worden (11 items HB); Persoonlijke inzet voor het schrijven van goede teksten (10 items PI).

De items werden in willekeurige volgorde in één vragenlijst aangeboden, die werd afgenomen in het cursusjaar, voorafgaand aan het hoofdonderzoek op 533 leerlingen uit derde klassen H.A.V.O. en V.W.O. van scholen die zich hadden aangemeld voor het hoofdonderzoek. De scores werden aan een statistische analyse onderworpen. De homogeniteit (Cronbachs alpha) van de schalen was redelijk tot goed: AK .90; BV .76; HB .84; HS .91; PI .71. Vijf items zouden op grond van een lage ($r \leq .30$) itemschaalcorrelatie geëlimineerd moeten worden. De correlaties tussen de schalen varieerden van .33 tot .73. De items van de schalen PI en BV correleerden met andere schalen vrijwel even hoog als met de eigen schaal.

Besloten werd de BV-schaal en de PI-schaal niet in het onderzoek te gebruiken, omdat we er niet in geslaagd waren voor deze variabelen schalen te construeren die zich onafhankelijk van andere schalen gedroegen. Daarmee verviel ook de mogelijkheid één schrijfangstscore samen te stellen door de drie angstcomponentschaalscores te sommeren. Van de andere schalen werden alleen die items gebruikt, die een correlatie vertoonden met de rest van de schaal die .15 groter was dan de correlaties tussen het item en een andere schaal. Zo ontstonden drie schalen: AK (14 items), HB (8 items) en HS (9 items). Op deze schalen zijn heranalyses uitgevoerd. De homogeniteit (Cronbachs alpha) van de schalen was bevredigend: AK .90; HB .80; HS .91. De itemschaalcorrelaties waren ≥ .47. Gezien de matige correlaties tussen de drie schalen (.41 - .48) mocht aangenomen worden dat deze schalen iets eigens representeerden. Kenmerkende items voor deze schalen waren: AK: 'Zelfs al moet ik nog met mijn opstel beginnen, denk ik al dat het me slecht af zal gaan'; HB: 'Ik wil graag dat mijn vrienden lezen wat ik geschreven heb'; HS: 'Schrijven is heel leuk'.

Dataverzameling. Op de drie toetsmomenten is aan alle leerlingen het instrument afgenomen, bestaande uit de 31 items die drie variabelen representeren. Om na te gaan hoe de schalen zich gedroegen in het hoofdonderzoek zijn ze nogmaals geanalyseerd. De items zijn zodanig gespiegeld dat een hoge score positief is: weinig angst (AK) of positieve attituden (HB en HS). De scores van proefpersonen die één of meer items geen score toekenden, zijn niet in de berekeningen betrokken.

Voor elke schaal zijn de homogeniteit (Cronbachs alpha), de item-restschaalcorrelatie en de correlaties tussen de somscores van items van de drie schalen berekend. Alle berekeningen zijn voor elke afname apart uitgevoerd. De correlaties tussen de schalen belopen waarden variërend van .27 tot .40, met een gemiddelde van .34. De schalen bleken homogeen. Voor de drie tijdstippen apart werden Cronbachs alpha en itemrestcorrelaties berekend. Voor de AK-schaal bedroegen de alpha's .90, .89

en nogmaals .89 en varieerden de itemrestcorrelaties van .35 tot .71. De alpha's van de HB-schaal waren .80, .81 en .82; de itemrestcorrelaties beliepen waarden van .29 tot .66. Voor de HS-schaal vonden we alpha's van .90, .93 en nogmaals .93 en itemrestcorrelaties, variërend van .56 tot .85.

## 5.5. Conclusies

Voor dit effectonderzoek konden we geen gebruik maken van bestaande meetinstrumenten. Een belangrijk onderdeel van het onderzoek was dan ook het ontwikkelen van instrumenten. Op basis van hetgeen in dit hoofdstuk is gerapporteerd, concluderen we dat we erin geslaagd zijn ook tot nu moeilijk meetbare Schrijfprestatievariabelen betrouwbaar te meten met behulp van opstelschalen. Voor Schrijfprocesvariabelen, een op het moment van ons onderzoek nog tamelijk onontgonnen terrein van onderzoek, werd een uitgebreid codeerschema ontwikkeld, waarmee objectief gecodeerd bleek te kunnen worden. Voor Psychologische Variabelen die aan schrijven gerelateerd zijn, werd een zelfrapportage-instrument ontwikkeld waarin drie interessante variabelen geoperationaliseerd werden. Problematisch is nog het instrument waarmee op analytische wijze getracht werd Doelgerichtheid en Publiekgerichtheid te meten. Opmerkelijk was dat de items uit het instrument voor Publiekgerichtheid tezamen geen betrouwbare test vormen. We vermoeden dat dit veroorzaakt wordt doordat de functie tussen de vaardigheid op Publiekgerichtheid en de scores op ons instrument niet lineair is. Ergens is er een optimum in het aantal items waarop men moet scoren: als men op 19 items positief scoort toont men wel dat men over een groot aantal technieken beschikt, maar tevens dat men het effect ervan op een lezer onjuist schat. Bij Doelgerichtheid bleek het instrument ook op itemniveau weinig betrouwbaar.

We concluderen dat de met behulp van de ontwikkelde instrumenten verkregen scores van voldoende kwaliteit zijn om te kunnen dienen voor de effectanalyses en dat we met behulp van de scores op de items van de analytische scoringsvoorschriften voor Doel- en Publiekgerichtheid uitsluitend exploratief, en dan nog op voorzichtige wijze, verschillen tussen scores van condities en toetsmomenten kunnen beschouwen.

# 6. METHODE EN PROCEDURES

Doel van het onderzoek was een experimentele steldidactiek te beproeven. Daartoe werd gebruik gemaakt van een experimenteel ontwerp met voor-, tussen- en natoets, waarin klassen geblokt werden op de variabele docent (Mellenbergh, 1977). We beschikten over acht scholen met in totaal elf docenten. Elke docent gaf les aan twee derde klassen, die aselect aan één van de twee condities werden toegewezen. In de experimentele conditie onderwezen de leerlingen elkaar door elkaars teksten schriftelijk te becommentariëren. In de controleconditie was de commentaartaak aan de docent voorbehouden. Afhankelijke variabelen waren Schrijfprestatievariabelen (Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl), Psychologische Variabelen (Angst-voor-het-niet-Kunnen-schrijven, Houding-jegens-Schrijven, Houding-jegens-Beoordeeld worden) en Schrijfprocesvariabelen. De vergelijking werd gemaakt met de klas als analyseniveau. Om mogelijke initiële verschillen tussen klassen te verdisconteren in de effectanalyses, werd zo mogelijk gebruik gemaakt van covariantie-analyses met de voortoetsscores als covariaat. Beschrijvende statistieken werden gebruikt om de kwaliteit van de instrumenten te bepalen en om toetsprestaties te beschrijven. Relaties tussen Psychologische Variabelen en Schrijfprestatievariabelen werden beschreven met behulp van Pearson product-moment correlaties. Schrijfprocesvariabelen werden met behulp van loglineaire modelpassing geëxploreerd.

In dit hoofdstuk beschrijven we de procedures. In par. 6.1. wordt de wijze van dataverzamelen beschreven inclusief de toetsafnameprocedures. In par. 6.2. komen de scholen, proefpersonen en docenten aan bod. In par. 6.3. beschrijven we op basis van docentlogboeken de praktijk van de onderwijsprocedures. In par. 6.4. worden de resultaten van een implementatie-onderzoek gerapporteerd. Voor het overzicht van alle onderzoeksmaatregelen verwijzen we naar tabel 6.1.

De activiteiten tijdens het cursusjaar waarin de effecten van de experimentele onderwijsprocedure getoetst werden, kunnen in drie groepen onderscheiden worden: het voorzien in onderwijsprocedures, het afnemen van toetsen en het controleren van de onderwijsprocedures. In tabel 6.1. worden deze activiteiten ten behoeve van een totaaloverzicht in schema gebracht.

Bij dit schema horen twee toelichtingen. In de oorspronkelijke opzet was het de bedoeling de leerlingen driemaal dezelfde schrijfopdracht voor te leggen. Op advies van de betrokken docenten herzagen we dat plan en ontwierpen voor de natoets een andere schrijfopdracht. Er werd gekozen voor een ontwerp met een tussentoets om eventueel demotivatie-effecten op te vangen die zouden kunnen ontstaan aan het eind van zo'n

Tabel 6.1.: Overzicht van onderzoeksactiviteiten in het toetsingsjaar

| | Onderzoeksmaatregelen | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Afhankelijke variabelen | Voortoets | Toewijzing van klassen aan conditie via lot | Cursus 1 | Cursus 2 | Cursus 3 | Tussentoets | Cursus 4 | Natoets |
| a) Schrijfprestatie-variabelen | Opdracht "Stress" | | | | | als voortoets | | opdracht "Huiswerk" |
| b) Psychologische variabelen | Selfrapportageinstrument | | | | | als voortoets | | als voortoets |
| c) Schrijfprocesvariabelen | Opdracht "Kinderen ja kinderen nee" | | | | | opdracht "Alleen is maar alleen" | | |
| Controle maatregelen | | | | | | | | |
| a) docentenlogboek | | | x | x | x | | x | |
| b) leerlingenlogboek | | | x | x | x | | x | |
| c) 'time-on-task' observaties | | | | x | | | x | |
| d) leerlingenvragenlijst | | | | x | | | x | |
| Periode | aug-sept | sept | sept-okt | nov-dec | jan-feb | feb | ma-apr | mei |

intensieve stelcursus. Alle te behandelen leerstof was in de eerste drie cursussen aan bod gekomen, zodat de tussentoets als eerste natoets te beschouwen is. Een tweede overweging om een vroege en een late natoets te plannen was de kans op uitval. Immers, we konden, statistisch gezien, over maar weinig klassen beschikken. Uitval door bijvoorbeeld ziekte van docenten zou het statistisch ontwerp zeer verzwakken.

## 6.1. Dataverzameling en toetsprocedures

Tabel 6.2. is een overzicht van het aantal leerlingen per klas bij het begin van het cursusjaar met vermelding (tussen haakjes) van het aantal leerlingen dat aan alle toetsmomenten meedeed. Dit kleinere aantal vormde de groep waaruit per klas willekeurig twaalf leerlingen getrokken werden. Aan de hand van de scores van deze twaalf leerlingen werden klassescores berekend. De subgroep van 264 leerlingen is representatief voor de gehele groep leerlingen die aan het onderzoek meewerkten. De verschillen tussen beide groepen wat de gemiddelden op de drie Psychologische Variabelen betreft, variëren van .04 tot .90. Het grootste verschil is bij lange na niet statistisch significant.

Van elk van de willekeurig gekozen leerlingen werden de voor-, tussen- en natoetsopstellen diplomatisch uitgetypt om beoordelaarsproblemen te reduceren. Boven het opstel werd een identificatienummer aangebracht. Op geen enkele wijze was het een beoordelaar mogelijk te achterhalen uit welke klas of conditie een opstel stamde.

Tabel 6.2.: Overzicht van het aantal leerlingen per school per klas, met vermelding (tussen haakjes) van het aantal leerlingen dat op drie toetsmomenten een opstel schreef

| | | Conditie | |
|---|---|---|---|
| Schoolnr. | Docentnr. | Experimenteel | Controle |
| 1 | 1 | 29 (24) | 30 (28) |
| 1 | 2 | 24 (15) | 22 (17) |
| 2 | 3 | 32 (22) | 32 (27) |
| 2 | 4 | 17 (12) | 19 (15) |
| 3 | 5 | 23 (21) | 25 (24) |
| 4 | 6 | 31 (17) | 31 (25) |
| 5 | 7 | 22 (21) | 23 (20) |
| 6 | 8 | 24 (16) | 22 (16) |
| 7 | 9 | 23 (17) | 25 (17) |
| 8 | 10 | 25 (21) | 20 (16) |
| 8 | 11 | 26 (16) | 26 (19) |

Voor de analyses van de Psychologische Variabelen dienden dezelfde leerlingen als proefpersoon als die waren geselecteerd voor de Schrijfprestatievariabelen. Voor het onderzoek naar Schrijfprocesvariabelen meldden zich - soms op vrijwillige basis, soms na aandrang van de docent -, via de docenten twee leerlingen per klas, een jongen en een meisje. Van alle 44 aangemelde leerlingen werd een voortoets- en een natoetsprotocol verzameld. Per duo werden daarna aselect van een de beide protocollen uitgetypt, in proposities ingedeeld en gecodeerd.

De afnames van de toetsen beschrijven we hieronder kort. De voortoets werd afgenomen in de tweede of derde les van de eerste schoolweek. De docent fungeerde als proefleider. Hij ontving een handleiding met daarin letterlijk uit te spreken teksten. Als voorzorgsmaatregel voor het verkrijgen van onafhankelijke scores werden de tafels uit elkaar gezet. Na een introductie en een voorbeeld werd gedurende 10 minuten de attitudetoets afgenomen. De ingevulde exemplaren werden door een leerling opgehaald en in een antwoordenvelop gedaan. Het moest voor de leerlingen duidelijk zijn dat de scores niet door de docent ingezien zouden worden. In datzelfde lesuur werden daarna de opstelopdrachtboekjes uitgedeeld, met de instructie zich in de opdracht te verdiepen. De boekjes werden aan het eind van het lesuur opgehaald en aan het begin van het volgende lesuur weer uitgedeeld, met de instructie:

> 'Je krijgt nu twee lesuren om het opstel te schrijven. Je maakt kladaantekeningen, of een kladversie, dat moet je zelf weten hoe je het best een opstel op poten kunt zetten. Na een half uur waarschuw ik je: het is goed om dan ongeveer aan je net te beginnen. Alle aantekeningen, kladstukken en het net komen in het boekje. Denk eraan: Schrijf een leesbaar net! Schrijf overtuigend en publiekgericht. Begin met het invullen van de gegevens aan de voorzijde."

De boekjes werden aan het eind van het lesuur opgehaald en het volgende lesuur weer uitgedeeld: 'Schrijf dit lesuur je opstel af. Bedenk dat het goed leesbaar moet zijn. Schrijf in zwart. Schrijf overtuigend en publiekgericht. Praat niet meer.' De toetsboekjes werden aan het eind van dit lesuur opgehaald en in het bijzijn van de leerlingen in een envelop gedaan die werd dichtgeplakt. Leerlingen die nog niet klaar waren, mochten in een tussenuur of pauze de laatste hand aan hun opstel leggen.

In de introductieles had de docent reeds twee vrijwilligers gevraagd voor het hardopschrijven. Deze leerlingen werden vrijgesteld van enige lessen, zodat zij ruim de tijd hadden een hardopschrijfopdracht te vervullen. Voor het inrichten van de testsituatie volgden wij de adviezen van Swarts, Hayes en Flower (1980). Na een kennismaking met een lid van het onderzoeksteam dat als proefleider fungeerde, werd een schriftelijke instructie gelezen. Daarna werd een denkpuzzel voorgelegd die hardop uitgevoerd moest worden. Vervolgens kreeg de proefpersoon een stukje van een cassettebandje te horen, waarop een hardopschrijvende leerling te horen was. Het fragment was zó geconstrueerd, dat alle mogelijke processen en subprocessen erin voorkwamen. De proefpersoon probeerde tenslotte zelf een paar regels naar aanleiding van een opdracht hardop te schrijven. Dit fragment werd gezamenlijk teruggeluisterd. Al deze instructiemaatregelen vergden circa 20 minuten. Als de proefpersoon geen verduidelijkingen meer wenste, werd de opstelopdracht verstrekt. De proefleider bediende de geluidsapparatuur, hield aantekeningen bij en weerhield de proefpersoon van stiltes door een enkele aansporing. Men verbleef in dezelfde ruimte als de proefpersoon, op een hem zo weinig mogelijk storende wijze.

De afnames van de tussentoets verliepen onder nagenoeg dezelfde omstandigheden. Aangezien dezelfde opstelopdracht werd gebruikt als voor de voortoets, werden er nu twee lesuren uitgetrokken voor de attitudenvragenlijst en de opstelopdracht samen. Er werd toegelicht waarom dezelfde opstelopdracht werd gebruikt. De proefpersonen die de hardopschrijfopdracht vervulden, kregen nu slechts een korte herhaling van de instructie over de bedoeling ervan. Zij kregen 1½ uur de tijd. Voor het hardopschrijfonderzoek was dit toetsmoment de natoets.

De natoets gold de Psychologische Variabelen en Schrijfprestatievariabelen. De leerlingen kregen twee lesuren de tijd om de attitudenvragenlijstjes in te vullen en de nog niet eerder uitgevoerde opstelopdracht te maken.

## 6.2. Proefpersonen, docenten, scholen en formeren van condities

Het experiment werd uitgevoerd in derde klassen (V.W.O. en H.A.V.O.). De leerlingen waren 14 à 15 jaar oud. Bij de keuze

voor deze populatie speelden de volgende overwegingen een rol. Eén doelstelling van de experimentele cursus was publiekgericht schrijven. Voor deze vaardigheid is vereist dat de schrijver zich de effecten van de communicatie op de ontvanger kan voorstellen. Deze metapositie nemen leerlingen naar verwachting niet veel eerder dan op die leeftijd in (Flavell, 1977; Glucksberg et al., 1975; Shantz, 1975). Een tweede overweging gold de praktijk van derde klassen. In schooltypes als het H.A.V.O. en Atheneum hebben derde klassen een wat onduidelijk curriculum. Enerzijds wordt de onderbouw afgesloten, anderzijds is er de voorbereiding op de bovenbouw. De curricula zouden er niet onder lijden als in dat leerjaar de derde klas als een proeftuin zou worden gebruikt. Het verkrijgen van medewerking van docenten zou voor derde klassen waarschijnlijk niet problematisch zijn. Een derde overweging gold de equivalentie van groepen. In vierde klassen, waarin het onderzoek qua lesinhouden ook goed zou passen, zijn de groepen bij het vak Nederlands ongelijk qua samenstelling door vakkenpakketkeuze. In derde klassen is daarvan geen sprake.

De leerlingen die bij het onderzoek betrokken werden, waren afkomstig van acht verschillende brede scholengemeenschappen. Deze scholen zijn geworven via een oproep in het lerarenvakblad Levende Talen en via een wervingsbrief. Deze brief werd verstuurd aan de secties Nederlands van 119 scholen met een V.W.O.-afdeling. Dit waren de scholen, die niet behoorden tot de 40 scholen die al voor een ander stelvaardigheidsonderzoek waren aangeschreven, maar wel voorkwamen op een computerlijst van het Centraal Instituut voor Toetsontwikkeling. Van de aangeschreven secties bleken er 38 in principe bereid tot deelname aan ons onderzoek. Vertegenwoordigers van deze secties werden uitgenodigd voor een informatieve bijeenkomst. Van 13 secties nam een vertegenwoordiging aan een van zes bijeenkomsten deel. Acht secties en hun schoolleidingen zegden toe bereid te zijn met één of twee docenten deel te nemen aan het onderzoek. Als negende school werd de school bereid gevonden die als proefschool had gediend. Om roostertechnische redenen moest één school afvallen die nog wel deelnam aan de finale beproeving (par. 4.3.)

De schoolleidingen van de acht resterende scholen ontvingen een brief waarin het doel van het onderzoek werd toegelicht en gevraagd werd de lesroosters van de beide klassen van elke deelnemende docent zoveel mogelijk gelijk te laten lopen. De directies zegden toe hiernaar te streven.

Van een drietal scholen namen twee docenten aan het onderzoek deel. Tabel 6.3. is een overzicht van de deelnemende scholen, het aantal docenten per school en de klassegrootte aan het begin van het schooljaar 1984-1985.

De schoolgrootte varieerde van 850 tot 1300 leerlingen, met een gemiddelde van 1065. Drie scholen hadden een rooms-katholieke signatuur, twee waren openbaar en de overige drie waren een bijzonder neutrale school, een protestants-christelijke en

Tabel 6.3.: Overzicht van de deelnemende scholen met leerlingenaantal per klas per aug. 1984

| School | Afdeling | Aantal leerlingen per klas: Experimentele klas | Controle klas |
|---|---|---|---|
| Heerlen | V.W.O. | 29 | 30 |
| | H.A.V.O. | 24 | 22 |
| Dordrecht | V.W.O. | 32 | 32 |
| | H.A.V.O. | 17 | 19 |
| Tilburg I | V.W.O. | 23 | 25 |
| Nieuwegein | H.A.V.O./V.W.O. | 31 | 31 |
| Tilburg II | V.W.O. | 22 | 23 |
| Wassenaar | V.W.O. | 24 | 22 |
| Waddinxveen | V.W.O. | 23 | 25 |
| Zierikzee | V.W.O. | 25 | 20 |
| | H.A.V.O. | 26 | 26 |

een oecumenische samenwerkingsschool. De twee openbare scholen herbergden drie schoolsoorten (M.A.V.O., H.A.V.O. en V.W.O.), de andere scholen waren H.A.V.O.-V.W.O.- scholengemeenschappen. Op één school waren de derde klassen heterogeen ingedeeld. Op alle andere scholen waren er evenveel typen derde klassen als er schoolsoorten waren in de desbetreffende school.

De grootte van de sectie varieerde van vijf tot acht docenten. In alle secties waren lesmethoden voorgeschreven. Naast verschillende soorten bloemlezingen of leesboeken, werden gebruikt: Functioneel Nederlands (4x), Taal Vandaag (2x), Taalgoed (1x) en Opbouw (1x). In het jaar dat vooraf ging aan het cursusjaar werd in verschillende mate aan de hand van de methode gewerkt. Twee docenten op dezelfde school gebruikten de methode helemaal niet. De andere docenten gebruikten de methoden in tenminste de helft van het aantal lesuren (2x), of in driekwart à vijf-zesde van de lesuren. Twee docenten werkten in alle uren met de methode.

De aandacht die men gewoonlijk gaf aan stelonderwijs in de derde klas verschilde. Als de twee docenten die op de proefschool lesgaven niet in deze berekening betrokken worden, dan varieert de schatting van de docenten van een half lesuur tot vijf lesuren van een vierwekelijkse termijn (twaalf lesuren): gemiddeld bijna tweeëneenhalf van de twaalf lesuren. De betrokken docenten zeiden gewoonlijk vier à vijf opstellen in de derde klas te laten schrijven; één gaf slechts twee opdrachten per cursusjaar. Twee secties werkten al lang met Gericht Schrijven in de eindexamens; één school was daar niet lang geleden ook toe overgegaan. De andere vijf scholen hadden nooit iets van doen gehad met deze examenvorm.

De docenten waren overwegend dertigers en veertigers, variërend van 27 tot 48 jaar, met een gemiddelde van ruim 37

jaar. Eén van de docenten was een vrouw. Er waren zeven eerstegraders; vier hadden een tweedegraadsbevoegdheid en waren allen studerend voor het MO-B examen. De onderwijservaring varieerde van vier jaar en vier maanden tot 25 jaar, met een gemiddelde van ruim vijftien jaar. Negen docenten hadden een volledige baan van 28 of 29 lesuren. Eén docent gaf 23 uur les en één veertien. Tijdschriften op het gebied van vakdidactiek werden door vijf docenten gelezen (Levende Talen 5x en Moer 1x). Een specialistisch tijdschrift op het gebied van de taalbeheersing werd door één docent gelezen; de andere docenten waren alleen op taal- en/of letterkundige tijdschriften geabonneerd.

Samenvattend stellen we dat de betrokken docenten geroutineerde docenten waren. De aandacht die zij in derde klassen aan stelvaardigheid zeggen te geven, is heel verschillend qua tijdsbesteding; het gemiddelde komt overeen met het gemiddelde van een grotere steekproef. Uit Damhuis et al. (1983) blijken in de onderbouw gemiddeld twee à drie lesuren besteed te worden aan stelonderwijs.

Zodra bekend was aan welke twee klassen de elf docenten les zouden geven tijdens het onderzoeksjaar, werd geloot tot welke conditie een klas zou behoren. Deze beslissing werd de docenten ná de voortoets meegedeeld. Tijdens de voortoets waren de omstandigheden in alle 22 klassen dus hetzelfde.

## 6.3. De onderwijsprocedures in de praktijk

In deze paragraaf wordt uitgebreid beschreven hoe de twee onderwijsprocedures in dit onderzoek zijn uitgevoerd. We doen dat uitvoerig om een helder beeld te scheppen van het gegeven onderwijs. Het is belangrijk te weten of de betrokken docenten zich aan de handleiding hebben gehouden en vast te stellen hoe ernstig een eigenhandig ingrijpen voor het onderzoek was. (De werkwijze kan ook opgemaakt worden uit de docentenhandleiding die integraal is opgenomen in het bijlagenrapport 1.) De validiteit van het onderzoek zou ernstig aangetast geweest zijn als een docent bijvoorbeeld de procedures verwisselde: de ene keer in de ene klas de experimentele onderwijsprocedure, de andere keer in de andere klas. De beschrijving biedt de lezer een mogelijkheid te controleren of de onderwijsprocedure een adequate operationalisatie van leerlingenrespons is en of het onderwijs in de experimentele en controlegroep niet op meer dan de relevante variabelen verschilden.

De beschrijving van de vier cursussen baseren we op door docenten en leerlingen bijgehouden logboeken. Elke docent werd verzocht per cursus een voorgedrukt logboek bij te houden. Tevens werd men verzocht een korte beschrijving te geven van de lessen die gegeven werden tussen twee cursussen in. Dit om te controleren of er lesactiviteiten ontplooid werden die interfereerden met het onderzoek. Hierop komen we in par. 6.4. terug.

De docent werd gevraagd erop toe te zien dat een klasselogboek bij werd gehouden tijdens de cursussen. Vóórin het schrift werd een klasselijstje geplakt. Leerlingen zorgden ervoor dat het schrift via dat lijstje rouleerde. Hun opdracht, ook voorin het schrift geplakt, was één pagina lang de les te beschrijven, het opgegeven huiswerk voor de volgende les te noteren, en een omschrijving van de werksfeer te geven. We ontvingen van acht docenten alle logboeken. Van een negende docent ontvingen we alleen zijn logboek over de eerste schrijfronde. De tiende docent stuurde alleen de leerlingenlogboeken in, die niet geheel volledig waren. De elfde docent stuurde, ook na aandringen en aanmanen, niets in. De beschrijving van de cursussen die hieronder volgt, is gebaseerd op de beschikbare docentlogboeken. De leerlinglogboeken zijn gebruikt als controle.

De beschrijvingen zijn als volgt opgebouwd. Eerst wordt beschreven welke onderwijsleeractiviteiten er successievelijk aan bod gekomen zijn. Daarna wordt ingegaan op ons bekende afwijkende lesuitvoeringen.

De eerste cursus bestond uit vijf voorbereidende lessen, twee commentaarlessen (alleen de experimentele groep), twee commentaarverwerklessen en een herschrijfles en werd gegeven in de periode tussen het begin van het cursusjaar (eind augustus, begin september) en de herfstvakantie (half oktober). Eén docent kon pas eind september beginnen en beëindigde de schrijfronde na de herfstvakantie. Alle docenten gaven de schrijflessen in één door andersoortige lessen ononderbroken periode van vier lesweken.

De voorbereidende lessen: drie theorielessen en twee schrijflessen. In de eerste les werd de werkwijze voor het gehele cursusjaar geïntroduceerd via een bordschema, het lespakket voor de eerste schrijfronde uitgedeeld en in multomappen ingedeeld en het huiswerk besproken. Dat bestond uit het bestuderen van de instructieteksten over doel- en publiekgerichtheid en het bestuderen van de schrijfopdracht waarover een week later geschreven moest worden. Om leerlingen vast te kunnen laten stellen of ze de instructieteksten begrepen, werden zelfcontrolevragen uitgereikt. Het bespreken van de antwoorden op de zelfcontrolevragen besloeg de tweede voorbereidende les. Als huiswerk werd opgegeven de instructieteksten over de reflectieopdrachten en de problemen van het commentaar op opstellen te bestuderen. Wederom werden controlevragen uitgereikt. In de derde les werden deze vragen besproken. Deze bespreking verliep niet zo intensief als de vragen over doel- en publiekgerichtheid. Leerlingen klaagden over het vele huiswerk, docenten over nogal wat leerlingen die hun huiswerk niet gemaakt hadden. Via de bespreking kwamen de hoofdpunten van het opgedragen werk toch onder de aandacht. Voorts schreven leerlingen een zelfportret, waarin hun schrijfervaring beschreven werd: 'Hoe schrijf ik en wat zijn mijn problemen?'. Als huiswerk werden opgegeven het in het net

schrijven van de antwoorden op de zelfcontrolevragen en het nogmaals bestuderen van de schrijfopdracht. De volgende twee lessen waren schrijflessen. In de eerste les werd een klad, een ontwerp of werden aantekeningen gemaakt en werd gestart met het net voor de eerste versie, die het volgende lesuur afgeschreven werd. Als leerlingen niet klaar waren, werd in een pauze of tussenuur doorgeschreven. De leerlingen in de docentgroep schreven op multopapier. De leerlingen in de leerlingengroep schreven op zelfdoorschrijvend papier, waardoor de docent in één keer over vier Kopieën beschikte. Als huiswerk werd opgegeven de eerste reflectieopdracht, die leerlingen opdraagt te beschrijven hoe het opstelschrijven dit keer verliep en te oordelen over hun ingeleverde eerste versie.

De commentaarfase: twee lesuren. In de docentgroepen besteedde men twee lesuren aan andere lesactiviteiten terwijl de opstellen thuis door de docent werden becommentarieerd op dezelfde wijze als in de leerlingengroep. Deze lesactiviteiten hadden op uitdrukkelijk verzoek van de projectleiding geen betrekking op stelonderwijs; via docentlogboeken konden we achteraf vaststellen dat docenten zich aan deze afspraak hielden. Alle docenten gelukte het de nakijktijd (20 minuten per opstel) tot één week te beperken. Voor de leerlingengroep verdeelde de docent de opstellen. Hiervoor was een verdeelschema opgesteld, waardoor elke commentator elke schrijfronde opstellen van andere leerlingen kreeg. In de eerste commentaarles werd de werkwijze nog eens kort beschreven en de leerlingen voorgedaan hoe zij hun papieren moesten ordenen. Zij hadden acht vellen met opsteltekst: het opstel van zichzelf en drie opstellen van twee kantjes van andere leerlingen. Elk opstel had een andere kleur: de setjes zelfdoorschrijvend papier waren samengesteld uit vier verschillende kleuren. De slechtst doorgedrukte kopie was voor de schrijver zelf. Alle leerlingen becommentarieerden de opstellen in een vaste kleurvolgorde. Op die wijze was gegarandeerd dat elke schrijver door één commentator als eerste becommentarieerd werd en niet, door bijvoorbeeld een onaantrekkelijk handschrift, door drie commentatoren als laatste zou worden bekeken. In de eerste commentaarles werd voor alle drie opstellen het reactieformulier ingevuld: een omschrijving van de eerste indruk in woorden en een reactie op twaalf stellingen. Daarna werd de tweede reflectieopdracht gemaakt: 'Wat heb ik geleerd door het lezen van drie opstellen?' Als huiswerk werd opgegeven voor het blauwe opstel het commentaarformulier in te vullen. In de volgende les werd kort ingegaan op problemen van leerlingen met het opgedragen huiswerk. Vervolgens werden de twee overige opstellen becommentarieerd. De derde reflectieopdracht was ook lesstof: 'Wat heb ik geleerd door het becommentariëren van drie opstellen?' Aangezien de opstellen het volgende lesuur aan de schrijvers teruggegeven moesten worden en de docent als voorbereiding de opstellen moest herverdelen, werd het becommentariëren afgemaakt in die les, of anders diezelfde dag in een pauze.

Commentaar verwerken: twee lesuren. Dit gebeurde in vier stappen: ordenen, samenvatten, kiezen en een herschrijfplan opstellen. Leerlingen deelden de tijd zelf in. Voor het

ordenen van het commentaar beschikten de leerlingen in de experimentele groep over een schema waarin zij de scores van de drie commentatoren op het reactie- en het commentaarformulier overzichtelijk bijeenbrachten. Het commentaar werd per categorie (doelgerichtheid, publiekgerichtheid, diversen/algemene opmerkingen) samengevat. De leerling schreef daarna een commentaar op de commentaren, waarin hij aangaf welk commentaar hij aanvaardde en welk hij afwees, en de keuze beargumenteerde. Tenslotte stelde hij een herschrijfplan op: een tweekolommenschema met in de linkerkolom wat de leerling wilde veranderen en in de rechterkolom hoe hij dit doel wilde verwezenlijken.

Herschrijffase: één lesuur. Beide groepen schreven op basis van hun herschrijfplan een definitieve versie. De leerlingen in de controlegroep maakten ter afronding hun tweede reflectieopdracht: 'Wat heb ik van deze schrijfronde geleerd?'

De leerlingen leverden hun mappen één van de volgende lesuren bij de docent in om deze te laten nakijken op compleetheid, orde en kwaliteit. Leerlingen kregen feedback daarover via een formulier. Ze konden hun mappen daarna herstellen voor een cijfer voor het eerste rapport.

Evaluatie. Een paar docenten hebben dit lesschema door omstandigheden iets gewijzigd. Twee docenten besteedden in de voorbereidende fase en in de fase waarin het commentaar verwerkt werd een lesuur extra aan de schrijfronde, in beide groepen. Een derde docent gaf zijn leerlingen maar één lesuur om het commentaar te verwerken tot een herschrijfplan, ook in beide groepen. Noch de 'overschrijding' noch de bekorting van de geplande lestijd beschouwden we als een afwijking van het onderwijsprogramma: alle dragende elementen daarvan zijn uitgevoerd.

Aan de tweede cursus lag dezelfde fasering ten grondslag als aan de eerste schrijfronde. Op basis van een evaluatieve vergadering met de betrokken docenten over de eerste schrijfronde werden enkele ingrepen verricht. Het was docenten gebleken dat leerlingen zich niet genoeg bewust waren van de fictieve communicatieve situatie van de schrijfopdracht. Ten tweede was ervaren dat de te bestuderen theorie over doel- en publiekgerichtheid te los van de te schrijven tekst werd gepresenteerd: er werd aanbevolen een meer geïntegreerde vorm na te streven. Ten derde vond men de belasting qua huiswerk in de schrijfronde ongelijkmatig. Aan deze drie wensen zijn we met enkele omzettingen tegemoet gekomen. Deze schrijfronde werd gegeven in de periode tussen begin november en half december en besloeg vier lesweken. De docent die de eerste schrijfronde later afsloot, begon pas eind november met de tweede ronde, die hij half januari beëindigde. Alle docenten werkten met hun beide groepen gelijk op. Niet alle docenten lukte het aan de tweede ronde een ononderbroken serie lessen te besteden: vanaf half november vielen lessen uit door rapportbesprekingen en gemeenschappelijke proefwerken.

De voorbereidende lessen: zes lesuren. Aan het einde van een 'gewone' les werd het nieuwe lesmateriaal uitgedeeld en een huiswerkopdracht verstrekt: het bestuderen van de schrijfopdracht en het bewerken met codes van de informatie over het onderwerp. De tweede voorbereidende les werd besteed aan een klasseleergesprek over de kenmerken van de opdrachtsituatie, (onderwerp, doel, publiek, medium, voorgeschiedenis) en aan het verwerken van de huiswerkopdracht via groepswerk, een onderwijsleergesprek of een individuele opdracht. Sommige docenten gaven de individuele opdracht bovendien als huiswerk op. Deze droeg de leerlingen op hun mening over de in de opdracht geformuleerde stelling te noteren, drie argumenten erbij te bedenken, benevens een ondersteunend citaat bij elk argument. Ditzelfde moest ook gedaan worden voor de tegenovergestelde mening. Het huiswerk voor de volgende les was nogmaals de schrijfopdracht te bestuderen. De volgende les schreven leerlingen het klad voor het middenstuk van hun opstel. Als huiswerk werd opgegeven de instructieteksten over structuur te bestuderen en de bijbehorende controlevragen te maken. In de daaropvolgende les werden de antwoorden klassikaal besproken en werd een individuele opdracht verstrekt: het controleren van het klad op structuur en het benoemen van alineafuncties. Het huiswerk voor de volgende keer was deze opdracht af te maken, instructieteksten over betrouwbaarheid te bestuderen en vier controlevragen hierover te beantwoorden. Deze werden in de volgende les besproken, waarna leerlingen individueel hun klad op betrouwbaarheid controleerden. Het huiswerk bestond uit het in het net schrijven van de antwoorden op de controlevragen en het nogmaals doorlezen van de schrijfopdracht. In de les daaropvolgend werd de eerste versie van het opstel geschreven en bij de docent ingeleverd.

Commentaar geven: twee lesuren. Er werden nu 20 commentaarvragen gebruikt. Uit de logboeken bleek dat het drie docenten niet gelukte de opstellen uit de controlegroep na een week weer terug te geven. Eén rondde de schrijfperiode net voor de kerstvakantie af en gaf na de vakantie de opstellen terug; ook de leerlingen in zijn experimentele groep kregen pas na de vakantie feedback over hun prestaties. Bij een andere docent viel een aantal lessen uit waardoor beide groepen twee weken op hun opstellen moesten wachten. Een derde docent had drie weken nodig voor de dertig opstellen uit zijn controlegroep, waardoor de leerlingen hun opstellen geruime tijd later terug kregen dan de leerlingen in de parallelklas. Een andere inbreuk op de planning maakte een docent die de opstellen uit de experimentele groep niet door drie, maar door twee leerlingen liet becommentariëren, wegens het uitvallen van lesuren door rapportvergaderingen.

Commentaar verwerken: twee lesuren. In het eerste lesuur werden de becommentarieerde opstellen teruggegeven en een werkplanning afgesproken, opdat leerlingen een week daarna in de les hun herschrijfplannen zouden kunnen opstellen, het tweede lesuur. De docenten merkten in hun logboeken op, dat de

les waarin zij het commentaar aan de leerlingen in de controlegroep uitdeelden, nogal onrustig was. Veel leerlingen vroegen om verduidelijkingen en uitleg over de symbolen en de nummers waarvan hun opstel door de docent voorzien was. Hieruit blijkt dat deze leerlingen de kennis misten die de leerlingen in de andere conditie wel hadden door het uitoefenen van de commentaartaak. Wellicht is er via de mondelinge toelichting van de docent extra instructie gegeven.

Herschrijven: één lesuur, als in eerste cursus.

Evaluatie. Voor zover wij na konden gaan hebben alle docenten zich aan het lessenplan gehouden. Eén docent besteedde één lesuur meer aan de voorbereidende lessen, - het doorwerken van de fragmenten -, maar deed dat in zijn beide groepen. Omdat de effectieve tijd nogal bekort werd door veel lokaalwisselingen, laste deze docent ook in de fase van commentaar verwerken een extra les in de docentgroep in. Een andere docent liet elk opstel in de leerlingengroep maar door twee leerlingen becommentariëren.

De derde cursus. De fasering van deze ronde was als de vorige. Alle docenten gaven hun introductieles vlak voor de kerstvakantie. Zeven docenten startten de eigenlijke cursus meteen na de kerstvakantie en beëindigden deze in de eerste week van februari. Eén docent kon pas eind januari beginnen omdat zijn tweede cursus tot in januari doorliep: hij beëindigde de derde half februari. Van de overige docenten weten we alleen dat zij de cursus begin februari afrondden, omdat we omstreeks die tijd afspraken maakten voor de afname van hardopschrijfprotocollen, die niet eerder plaatsvond dan ná de derde ronde.

Voorbereidende lessen: vijf lesuren. In de eerste les werd de nieuwe schrijfopdracht uitgedeeld. Bij de schrijfopdracht was een aantal zogenaamde voorbereidende oefeningen opgenomen, die de leerlingen gedeeltelijk in de kerstvakantie moesten maken. Daaronder waren het verrichten van een klein onderzoekje om nieuwe informatie te genereren, het analyseren van de opdrachtsituatie en het formuleren van het eigen en het tegenovergestelde standpunt.

In de eerste les na de vakantie kregen de groepjes gelegenheid hun onderzoeksgegevens uit te wisselen en te tabelleren. Daarna leidde de docent een klassegesprek over het onderwerp. Als huiswerk werd opgegeven de instructieteksten over nieuwswaarde en taalgebruik te bestuderen en het tweekolommenschema waarin argumentatie en contra-argumentatie geschematiseerd werd te vervolmaken op grond van de bespreking in de klas. In vrijwel alle klassen waren er leerlingen die hun kersthuiswerk niet gemaakt hadden, zodat hun groep informatie voor het onderzoekje moest ontberen. Deze leerlingen moesten die onderzoeksgegevens alsnog produceren. In het volgende lesuur werden de zelfcontrolevragen besproken over de instructieteksten. Er kon begonnen worden aan het huiswerk: de antwoorden op de zelfcontrolevragen in het net schrijven en de schrijfopdracht bestuderen.

Voor het overige verliep deze schrijfronde als de eerste ronde: twee lessen schrijven (een klad en een net), twee lessen commentaar geven (alleen in de experimentele groep uiteraard), twee lessen commentaar verwerken en één les herschrijven. Over het commentaar geven moet nog vermeld worden dat nu niet alle beschikbare dertig commentaarvragen gebruikt werden, maar een selectie daaruit. De vijf vragen van elke nieuwe categorie, nieuwswaarde en taalgebruik, waren daarin opgenomen en verder twee vragen per 'oude' categorie (doelgerichtheid, publiekgerichtheid, structuur en betrouwbaarheid). De tijd die docenten nodig hadden om de opstellen becommentarieerd terug te geven, varieerde van een tot twee weken. De docenten die er langer dan een week over deden, hadden te kampen met ziekte (2x) waardoor lessen uitvielen en met extra werk door de overname van een conrectorsfunctie (1x).

Evaluatie. Van dit lessenplan werd een paar keer afgeweken. Eén docent liet leerlingen (per ongeluk) toch alle dertig commentaarvragen gebruiken, terwijl hijzelf wel een selectie maakte toen hij de opstellen uit de controlegroep van commentaar voorzag. Een tweede docent had maar één lesuur in de experimentele groep ter beschikking in de fase van commentaar geven, omdat er plotseling een lesuur uitviel. De leerlingen moesten daardoor twee opstellen thuis becommentariëren. Verder waren er twee docenten die in hun beide groepen in plaats van twee slechts één lesuur besteedden aan het verwerken van het commentaar. Zij lieten de leerlingen het herschrijfplan thuis maken. Een andere docent gaf de leerlingen in beide groepen juist een uur extra voor de fase van verwerken en herschrijven.

Deze overtredingen zijn niet zo ernstig: alle docenten van wie wij gegevens hadden werkten met hun groepen de geplande activiteiten af. Ernstiger was dat een docent zijn lestaak moest opgeven ten faveure van een conrectorschap. Twee jonge, pas afgestudeerde eerstegraads docenten, namen ieder een klas van hem over. Het experimenteel ontwerp met de docent als blokvariabele was daarmee verstoord. Omdat de eigenlijke docent verreweg het grootste gedeelte van de cursussen onderwees, beschouwen we het ontwerp tot en met de derde schrijfronde als ongeschonden.

De vierde cursus. Synthese. De laatste lessenserie had een samenvattend karakter. Het doel was de in de drie vorige schrijfronden verspreide procesaanwijzingen te integreren tot een schrijfaanpak in tien stappen. Het voorbereiden van een te schrijven tekst met een plan, tot nu toe geoefend met een herschrijfplan, werd nu nadrukkelijk vóórin het schrijfproces geplaatst. De nadenkoefeningen (reflectie op de eerste versie, het noteren van leerervaringen tijdens het lezen en becommentariëren van andermans opstellen) waren nu geïntegreerd in het stappenplan. Het doel van deze notities was nu niet meer het innemen van een metapositie om zo meer vat te krijgen op het

onderwijsleerproces, maar het verbeteren van de eerste versie van het opstel. Al dergelijke notities waren bouwstenen voor het herschrijfplan. De cursusduur was drie weken. De docenten konden kiezen uit twee varianten. In de ene variant bereidden leerlingen hun eerste versie thuis voor (huiswerk voor één week), in de andere kregen leerlingen drie lesuren gelegenheid een gedeelte van de voorbereidende opdrachten uit te voeren. Zes docenten kozen ervoor de leerlingen in de les te laten werken. Drie docenten lieten hun leerlingen de eerste versie thuis voorbereiden. De periode waarin de vierde schrijfronde gegeven werd, varieerde van twee maanden tot drie weken. De periode werd gekenmerkt door verbrokkeling door mondelinge schoolonderzoeken, paasvakantie en andere vrije dagen in april en mei. Eén docent gaf de cursus in de drie weken die aan de paasvakantie (begin april) vooraf gingen. Twee docenten begonnen na de paasvakantie (half april). Zes docenten gaven de introductieles vóór de paasvakantie en begonnen met de cursus meteen na de vakantie. Alle docenten planden hun beide groepen parallel qua tijd; ook de twee docenten die elk één groep hadden overgenomen werkten gelijk op.

Voorbereidende opdrachten: drie lesuren. Ze omvatten zes van de tien stappen. Na een opdracht waarin de schrijfopdrachtsituatie werd geanalyseerd, het tweekolommenschema dat al vaker gebruikt was, een schrijfplan en een 'denkversie', werden de leerlingen ertoe gebracht de versie zelf te controleren. Zij moesten daartoe verscheidene stukken uit de instructieteksten herlezen. Vervolgens schreven zij op school hun eerste versie die zij bij de docent inleverden.

Commentaarlessen: twee lesuren. Deze taak was ten opzichte van de vorige cursussen veranderd. Het reageren als natuurlijke lezer via het reactieformulier met de twaalf stellingen was vervallen, omdat leerlingen te kennen hadden gegeven dat tijdens het lezen van de opstellen de commentaarvragen al in het hoofd speelden. Alleen de extra opdrachten bij doel- en publiekgerichtheid werden gemaakt, er werden twee commentaarvragen per categorie gebruikt en er werd een eerste indruk in woorden geschreven.

De docenten hadden vier tot dertien dagen nodig om de opstellen van de leerlingen uit de controlegroep terug te geven. Nog steeds rapporteerden docenten dat er in die klassen in het lesuur waarin de opstellen werden teruggegeven veel toelichtingen werden gevraagd.

Commentaar verwerken: twee lesuren als in vorige cursussen.
Herschrijven: één lesuur, als in vorige cursussen.

Evaluatie. Er waren twee docenten die enigszins van het lessenplan afweken. Beiden besteedden zij maar één lesuur aan het verwerken van het commentaar. Eén van beiden liet de leerlingen in de leerlingengroep geen nummers meer boven de opstellen schrijven, maar stond toe dat zij hun namen gebruikten als identificatie zo zij dat verkozen. Het was duidelijk geworden dat de anonimiteit niet echt meer nodig was. Bovendien werkte de truc ook niet meer omdat handschriften herkend werden.

Conclusie. Vrij uitvoerig hebben we beschreven hoe de twee onderwijsprogramma's in de praktijk van een aantal docenten ten uitvoer werden gebracht. We hebben dat gedaan om te laten zien welke de cursussen waren waarvan de resultaten vergeleken zullen worden. Bovendien was het nodig in te gaan op de 'overtredingen' die docenten maakten tegen het voorgeschreven lessenplan. Onze conclusie op basis van de ons bekende gegevens is, dat de cursussen per docent parallel zijn uitgevoerd en dat verschillen tussen de lesplannen van de docenten niet zodanig zijn, dat we die zouden aanwenden om verschillen in effecten tussen groepen binnen docenten te verklaren.

Het is jammer dat het ons niet gelukt is van alle docenten logboeken te ontvangen. Of we de lessen van de drie docenten waarvan we geen of geen volledige serie logboeken ontvingen ook tot uitvoeringen van de voorgeschreven lessen mogen rekenen, is niet zeker. Andere bronnen (observaties, vragenlijsten) staan ons ten dienste om daarover meer zekerheid te krijgen.

Op één school moest een docent zich terugtrekken uit het project. Hoewel uit de logboeken blijkt dat de twee docenten die hem vervingen parallel werkten, is daarmee toch het experimenteel ontwerp verstoord voor wat betreft de vierde schrijfronde.

## 6.4. Controles op de uitvoering van de onderwijsprocedures

In de vorige paragraaf is uitvoerig beschreven hoe de twee onderwijsprocedures zijn uitgevoerd. Er trad hier en daar aan het licht dat er verschillen bestonden in het uitvoeren van de onderwijsprogramma's tussen docenten. Op nog twee andere wijzen hebben we gegevens verzameld over de implementatie van de onderwijsprogramma's, namelijk via observaties en vragenlijstonderzoek. Resultaten van deze twee controlerende activiteiten worden hierna gepresenteerd. Daarvoor gaan we in op het voorkomen van andere onderwijsactiviteiten die verantwoordelijk gesteld zouden kunnen worden voor eventueel optredende effecten op de afhankelijke variabelen.

Rivaliserende lesactiviteiten. Om te voorkomen dat docenten naast de twee onderwijsprogramma's ook eigen schrijfonderwijs zouden geven, werd de docenten gevraagd in de overige lessen vóór de natoets in mei geen lessen te geven waarin het stellen beoefend werd. Alle docenten zegden toe dit niet te zullen doen en zeiden na afloop van het cursusjaar dit ook niet gedaan te hebben. Men had trouwens de weinige lestijd die over was hard nodig voor andere onderdelen van het vak Nederlands! Om enig inzicht te verkrijgen in de lessen die men gaf, werd ook van de lesuren buiten de cursus een kort verslag gevraagd. Van negen docenten ontvingen we korte lesverslagen over de lesperioden tussen de schrijfcursussen. In deze lessen probeerden de docenten hun normale derde klasprogramma uit te voeren, voorzover de beperkte tijd dat toeliet. In alle

gevallen bleken de docenten parallel te werken: als er al sprake zou zijn van rivaliserende lesactiviteiten dan zijn beide klassen van een docent aan die invloed blootgesteld geweest. Of er sprake zou kunnen zijn van dergelijke lesactiviteiten, is moeilijk vast te stellen, omdat gegist moet worden wat bijvoorbeeld de invloed van zinsontleden is op het produceren van teksten. In andere klassen werd veel tijd aan het lezen van fictie besteed en aan stijlfiguren en -fouten. Enige invloed hiervan op de stijl in opstellen is te verwachten. Veel docenten lieten leerlingen regelmatig een tekstverklaring maken: invloed hiervan op de opbouw van opstellen is niet uitgesloten. In elk geval is er geen onderwijs in het schrijven van opstellen gegeven naast het onderzoeksprogramma. Mochten er verschillen geconstateerd kunnen worden tussen de twee condities in dit onderzoek, dan zijn deze niet aan verschillen in overige lesactiviteiten te danken: men gaf steeds in beide klassen hetzelfde.

Observaties. Een van de wijzen waarop men de implementatie van een onderwijsprogramma kan vaststellen, is na te gaan of en in welke mate leerlingen in de les bedoelde activiteiten uitvoeren. Hiertoe is een implementatie-instrument ontwikkeld. Omdat de onderwijsprogramma's beide tot in detail zijn voorgeschreven, konden we een gespecificeerde lijst opstellen van te realiseren lesactiviteiten, daarbij geïnspireerd door het door Van de Grift (te verschijnen) besproken onderzoek van Yeager, Wood en Kissel (1970). Er werden vier hoofdcategorieën onderscheiden: individuele activiteiten, docent-leerling activiteiten, leerling-leerling activiteiten en niet-lesgebonden activiteiten. Elke hoofdcategorie bevatte 7 à 10 activiteiten. De opdracht voor de observator was de activiteiten van vier in het leslokaal verspreid zittende leerlingen te scoren. In acht tijdsblokken van vijf minuten werd een leerling driemaal per minuut gescoord. Na vier minuten volgde een pauze van één minuut. Tevens noteerde de observator de beginsituatie in de klas en de afronding van de les.

Tijdens de tweede en de vierde cursus werden lessen geobserveerd. In de tweede cursus werden achttien klassen bezocht door in totaal vier observatoren. Er werden vergelijkbare lessen geobserveerd die alle de voorbereiding tot de schrijfopdracht als leerstof hadden. In de vierde ronde werden weer achttien klassen bezocht, nu door één observator. De lessen die geobserveerd werden, hadden alle betrekking op één van de laatste stappen in het onderwijsprogramma: het verwerken van commentaar. De klassen die in de tweede en vierde cursus uitvielen, waren niet dezelfde.

Van elke geobserveerde leerling is vastgesteld welk percentage van de lestijd (max. 40 minuten) deze zich bezighield met een bedoelde lesactiviteit. Over de beide groepen werd het gemiddelde berekend. Het verschil werd getoetst met de t-toets voor onafhankelijke waarnemingen (zie tabel 6.4.).

Tabel 6.4.: De gemiddelde lesinhoudgebonden tijdsinvestering van de twee condities in percentages (n=36) (E=experimentele groep; C=controlegroep)

| Moment van observeren | Groepen E | | Groepen C | | |
|---|---|---|---|---|---|
| | gem. | sd | gem. | sd | t-waarde |
| 2e cursus | 83.80 | 17.24 | 72.14 | 17.27 | 2.87* |
| 4e cursus | 78.72 | 16.27 | 73.56 | 15.73 | 1.36 |

* significant $p \leq .05$

Uit tabel 6.4. is op te maken dat de gemiddelde leerling om en nabij dertig minuten, driekwart van de lestijd, die dingen doet die van hem verlangd werden. Genoeg om te kunnen stellen dat zij met ons programma hebben gewerkt. De experimentele groep is intensiever bij het programma betrokken dan de controlegroep. Wij concluderen dat, gezien de observatiegegevens, de leerlingen in voldoende mate aan de onderwijsprogramma's hebben meegedaan om ervan te kunnen profiteren.

Vragenlijsten. Na de tweede en de vierde schrijfcursus werd de leerlingen een vragenlijst voorgelegd. Met deze vragenlijst werden verschillende doelen beoogd. Allereerst werd nagegaan of volgens de leerlingen alle geplande lesactiviteiten hadden plaatsgevonden. Ten tweede kon nagegaan worden welke feedbackboodschappen de leerlingen ontvingen en konden eventuele verschillen tussen beide condities op dit punt vastgesteld worden. Ten derde konden we achterhalen wat leerlingen zelf van de kwaliteit van de feedback vonden die zij kregen en welke invloed het commentaar volgens hen had op hun herschreven versie van het opstel.

De vragenlijsten bestonden uit 61 à 86 vragen, afhankelijk van de schrijfcursus en de conditie. Getracht werd de vragen over de cursussen gelijkluidend te formuleren, opdat een vergelijking tussen de twee momenten mogelijk zou zijn. Helaas beschikten we niet over een volledig bestand aan gegevens: van één docent raakten de vragenlijsten over de tweede cursus postaal zoek. Beschikt werd over de volgende aantallen: Experimentele groep 240, respectievelijk 265; Controlegroep 225, respectievelijk 273. Per vraag is een frequentieverdeling gemaakt van de gegeven antwoorden per conditie. In het hierna volgende rapporteren we in tabellen het percentage leerlingen binnen een conditie dat zegt een opdracht niet uitgevoerd te hebben.

Omstandigheden. We stelden een aantal vragen om inzicht te krijgen in eventuele verschillen tussen de condities in de omstandigheden waarin leerlingen met het onderwijsprogramma werkten. We onderscheidden vier factoren die invloed zouden kunnen uitoefenen op het functioneren van beide curricula: de mate van zelfstandigheid van leerlingen (7 vragen), de door

leerlingen waargenomen docentinspanning (5 vragen), werksfeer (5 vragen) en inzet (8 vragen). In tabel 6.5. staan de gemiddelden van de percentages leerlingen die een vraag positief beantwoordden.

Tabel 6.5.: Omstandigheden waarin de onderwijsprogramma's functioneerden. Gemiddeld percentage leerlingen dat een aantal vragen over deze factoren positief beantwoordde (E=experimentele groep; C=controlegroep)

| | Cursus 2 | | Cursus 4 | |
|---|---|---|---|---|
| Factoren | E | C | E | C |
| Zelfstandigheid | 70 | 73 | 68 | 74 |
| Docentinspanning | 77 | 74 | 77 | 78 |
| Werksfeer | 64 | 66 | 61 | 74 |
| Eigen inzet | 59 | 56 | 55 | 58 |

De verschillen tussen beide groepen in beide curussen zijn gering. Alleen wat de werksfeer betreft in de vierde cursus blijken er in de controleconditie meer leerlingen te zijn die de sfeer positief waarderen. Op alle vragen in deze categorie antwoorden relatief meer leerlingen uit de controleconditie positief dan leerlingen uit de experimentele conditie.

Deelname aan voorbereidende en begeleidende lesactiviteiten. In tabel 6.6. zijn de voorbereidende en begeleidende lesactiviteiten van beide cursussen opgenomen. De volgorde van de activiteiten is niet die van het onderwijsprogramma: ter wille van de vergelijkbaarheid hebben wij vergelijkbare lesactiviteiten van beide cursussen naast elkaar gezet. In de tweede cursus zijn vrijwel alle voorbereidende lesactiviteiten en begeleidende opdrachten door een flink deel van de leerlingen gemaakt. Alleen het bespreken van de teksten over het onderwerp is in de ogen van veel leerlingen niet gebeurd. Uit de lesverslagen bleek echter dat er wel in de klas over het onderwerp gesproken is. De formulering in de vragenlijst ('bespreken van teksten en citaten') is wellicht niet valide: als er losjes in de klas over het onderwerp gesproken wordt, is dat niet opgevat als 'bespreken van'. We concluderen, dat beide cursussen gegeven zijn zoals we planden, dat de leerlingen in voldoende mate deelnamen aan het onderwijs en dat de verschillen tussen de twee condities niet zodanig zijn dat deze de oorzaak kunnen zijn van eventuele verschillen op de afhankelijke variabelen.

Tabel 6.6.: Deelname aan voorbereidende en begeleidende lesactiviteiten. Het percentage leerlingen dat zegt niet aan de activiteit te hebben deelgenomen (E=experimentele groep; C=Controlegroep)

| Cursus 2 | E | C | Cursus 4 | E | C |
|---|---|---|---|---|---|
| 1. Lezen en bewerken van teksten over het onderwerp | 8.4 | 5.5 | 1. Bestuderen van de informatie over het onderwerp | 3.9 | 2.2 |
| 2. Bestuderen van de leerteksten over structuur en betrouwbaarheid | 10.5 | 5.0 | 2. Bestuderen van de leerteksten over tien stappen | 14.7 | 14.3 |
| 3. Bestuderen van leerteksten over de reflectie-opdracht 'Reactie op eigen opstel' | 17.7 | 25.5 | 3. Herhalen van oude theorie | 48.5 | 42.1 |
| 4. Het maken van een zg. kolommenschema om standpunt en argumenten te genereren en te ordenen (facultatief) | 23.8 | 17.6 | 4. Formuleren van standpunt | 8.5 | 4.8 |
| 5. Schrijven eerste versie | 2.5 | 3.6 | 5. Schrijven eerste versie | 9.8 | 7.1 |
| 6. Analyse eigen eerste op structuur en betrouwbaarheid | 19.8 | 22.4 | 6a. Eerste versie controleren | 29.6 | 26.6 |
| | | | 6b. Eerste versie bewerken | 33.6 | 31.2 |
| 7. Schrijven van lezersversie | 2.1 | 1.4 | 7. Schrijven van lezersversie | 5.9 | 5.6 |
| 8. Uitvoeren van reflectie-opdracht 1 (zie 10) | 9.2 | 7.7 | 8. Eigen reactie noteren | 32.6 | 33.8 |
| 9. Uitvoeren van reflectie-opdracht 2: Wat geleerd door lezen van andere opstellen | 4.7 | nvt | 9/10. Tijdens becommentariëren aantekeningen maken voor herschrijfplan | 42.3 | nvt |
| 10. Uitvoeren reflectie-opdracht 3: Wat geleerd door becommentariëren van andere opstellen? | 7.0 | nvt | | | |
| 11. Uitvoeren van reflectie-opdracht 4: Wat geleerd van hele schrijfcursus? | nvt | 15.4 | 11. Afsluitende reflectie-opdracht | 44.7 | 41.7 |
| 12. Volgen van een uitleg door de docent over de schrijfopdracht | 15.7 | 10.9 | | | |
| 13. Bespreking van een aantal teksten over het schrijfonderwerp | 49.1 | 47.6 | | | |
| 14. Zelfcontrole opdrachten maken | 3.8 | 3.6 | | | |
| 15. Volgen van de behandeling van de zelfcontrole-opdrachten | 3.0 | 7.0 | | | |
| | | | 16. Analyse van de schrijfsituatie | 8.9 | 3.7 |
| | | | 17. Inhoud zoeken | 9.9 | 4.5 |
| | | | 18. Schrijfplan opstellen | 17.5 | 13.7 |

Deelname aan de commentaartaak. De leerlingen in de experimentele condities vervulden een aantal commentaartaken voor drie opstellen van medeleerlingen. Sommige taken waren in beide cursussen gelijk. Om een indruk te krijgen van de feedback die leerlingen gekregen zouden hebben, wordt in tabel 6.7. een overzicht gegeven van alle commentaartaken in beide

cursussen, met vermelding van het percentage leerlingen dat de taak voor minder dan drie opstellen verrichtte. Dit percentage is te beschouwen als een indicatie voor het deelnameverlies. We merken op dat uit de docentlogboeken over de tweede schrijfcursus bleek dat er in één klas een groot aantal leerlingen (75%) slechts twee opstellen te becommentariëren kreeg (zie par.6.3.). Uit het antwoordenpatroon op de desbetreffende vraag bleek er nog één klas te zijn, waarin de docent er niet in geslaagd was (bijna) alle leerlingen op drie opstellen commentaar te laten geven. In deze klas deed slechts tweederde van de leerlingen dit. In alle andere klassen was er nauwelijks een leerling die minder dan drie opstellen kreeg voorgelegd. Tijdens de tweede cursus gaf 85% van de leerlingen op drie opstellen commentaar, tijdens de vierde cursus was dat 97%.

Tabel 6.7.: Deelname van commentaartaken in de experimentele conditie. Het percentage leerlingen dat de activiteit niet verrichtte

| Commentaartaken | Cursus 2 | 4 |
|---|---|---|
| 1. Invullen van 12 reactiestellingen | 3.8 | nvt |
| 2. Omschrijving van de eerste indruk | 5.0 | 12.4 |
| 3. Beantwoorden van commentaarvragen | 3.9 | 7.1 |
| 4. Omschrijven van het doel van de schrijver | 8.4 | 10.6 |
| 5. Golflijntjes onder publiekgerichte zinnen | 36.7 | 29.8 |
| 6. Markeren van driedeling (inleiding, kern, slot) | 37.0 | nvt |
| 7. Markeren van hoofdzinnen in alinea's | 30.1 | nvt |
| 8. Markeren van verbindende woorden en zinnen | 49.1 | nvt |
| 9. Markeren van structuuraanduidende woorden en zinnen | 68.2 | nvt |
| 10. Reactietekens in de kantlijn plaatsen | 22.3 | nvt |
| 11. Verwijsnummers van commentaarvragen in opstel plaatsen | 40.8 | 40.1 |
| 12. Toelichtingen op antwoorden schrijven | 6.9 | 1.9 |
| 13. Opmerkingen in het opstel schrijven | 9.4 | 10.0 |

De kernonderdelen van de commentaartaak zijn de nummers 1,2,3,11,12,13. Deze zijn in sterke mate geïmplementeerd, behalve het noteren van verwijsnummers van commentaarvragen in het te becommentariëren opstel. Uit de gegevens over de tijdsbesteding van commentatoren (tabel 6.8.) blijkt dat deze zeer acceptabel is.

Tabel 6.8.: Tijdsbesteding van commentatoren in de experimentele conditie per opstel in percentages

| Tijdsbesteding | Cursus 2 | Cursus 4 |
|---|---|---|
| 15 minuten | 11.5 | 11.4 |
| 15 à 30 minuten | 54.9 | 43.3 |
| 30 à 45 minuten | 25.7 | 18.9 |
| 45 minuten | 8.0 | 23.6 |

In de cursus werd ernaar gestreefd de opstellen anoniem te laten becommentariëren. In de vierde ronde is dit principe vrij gelaten: wie wilde mocht zijn naam boven zijn opstel schrijven. Uit tabel 6.9. blijkt dat volstrekte anonimiteit ook in de tweede ronde niet bereikt is.

Tabel 6.9.: Anonimiteit van schrijvers (experimentele groep) van opstellen in percentages leerlingen dat zegt geen, één of meer schrijvers te kennen

| Aantal schrijvers bij de commentator bekend | Cursus 2 | Cursus 4 |
|---|---|---|
| 0 | 12.4 | 17.8 |
| 1 | 31.3 | 14.7 |
| 2 | 24.9 | 15.4 |
| 3 | 31.3 | 51.4 |

We stelden tenslotte een viertal vragen over de perceptie van de leerlingen van hun inzet bij het commentaar geven en het effect ervan. In de tweede cursus gaf 83.8% te kennen op twee of drie opstellen hun best gedaan te hebben; in de vierde cursus was dit 95%. De aandacht die men besteedde aan de taak nam toe (89.7% in cursus 2 besteedde veel aandacht, 94.9% in cursus 4). Over de duidelijkheid van het commentaar werden de commentatoren (nog) optimistischer: 90.5% (cursus 2) en 97.2% (cursus 4) achtte het gegeven commentaar duidelijk. In dezelfde orde lag het percentage dat verwachtte dat de schrijver wat aan hun commentaar had: 92.4% (cursus 2) en 96.3% (cursus 4).

We concluderen dat de commentaartaak sterk geïmplementeerd is.

De aard en hoeveelheid van het ontvangen commentaar. Zeggen dat je commentaar gaf, wil nog niet zeggen dat je het ook déêd. In een van de vooronderzoeken (par.4.3.) bleek al dat leerlingen in de rol van commentaarontvanger minder commentaar leken te ontvangen dan zij gaven. We legden hun de verscheidene commentaartaken voor en vroegen hoeveel van de drie commentatoren deze taak voor hun opstel hadden verricht. We

beschouwden de taak als geïmplementeerd als twee van de drie commentatoren de taak verricht hadden. Aan de leerlingen in de controleconditie legden we deze taken voor met de ja-nee-vraag of deze taak door hun docent verricht was voor hun opstel.

In de tweede cursus zijn van de kerntaken (1,2,3,11,12 en 13) in de experimentele conditie het noteren van verwijsnummers (11) en het maken van 'vrije' opmerkingen in het opstel onvoldoende geïmplementeerd. Als docenten commentaar geven, noteren ze wel verwijsnummers (11) en schrijven ze wel commentaar in het opstel (tabel 6.10.)

In de vierde cursus is de hoeveelheid leerlingen die te klagen heeft over de hoeveelheid commentaar geringer. Alleen de omschrijving van de eerste indruk (nr.2) is vaker overgeslagen. Toch is het aantal leerlingen dat zegt geen of weinig verwijsnummers te krijgen en 'vrije' opmerkingen moet ontberen nog groot (resp. 48 en 52%). In de controleconditie is weinig veranderd tussen de tweede en vierde cursus. We concluderen dat leerlingen in beide condities vooral de resultaten van de taken 1,2,3,4 als feedback kregen; de leerlingen in de controleconditie kregen daarnaast 'vrije' opmerkingen in het opstel van hun docent en verwijsnummers in het opstel. Waarschijnlijk is de aard van de feedback in beide condities verschillend.

Tabel 6.10.: De aard en hoeveelheid feedback in beide condities op twee momenten. Het percentage leerlingen per conditie dat zegt geen commentaar van de docent (controleconditie) of geen of weinig commentaar van één medeleerling (experimentele conditie) gekregen te hebben. (E=experimentele groep, C=controlegroep)

| Commentaartaken | Cursus 2 | | Cursus 4 | |
|---|---|---|---|---|
| | E | C | E | C |
| 1. Invullen van 12 reactiestellingen | 5.7 | 14.0 | nvt | nvt |
| 2. Omschrijving van de eerste indruk | 12.3 | 15.2 | 15.9 | 17.3 |
| 3. Beantwoorden van commentaarvragen | 10.3 | 18.6 | 9.1 | 19.3 |
| 4. Omschrijven van het doel van de schrijver | 8.4 | 9.5 | 11.3 | 13.9 |
| 5. Golflijntjes onder publiekgerichte zinnen | 54.3 | 26.9 | 36.4 | 39.4 |
| 6. Markeren van driedeling (inleiding, kern, slot) | 48.8 | 34.9 | nvt | nvt |
| 7. Markeren van hoofdzinnen in alinea's | 49.0 | 26.9 | nvt | nvt |
| 8. Markeren van verbindende woorden en zinnen | 64.7 | 30.6 | nvt | nvt |
| 9. Markeren van structuuraanduidende woorden en zinnen | 77.7 | 53.7 | nvt | nvt |
| 10. Reactietekens in de kantlijn plaatsen | 44.4 | 33.0 | nvt | nvt |
| 11. Verwijsnummers van commentaarvragen in opstel plaatsen | 61.1 | 23.3 | 48.2 | 23.3 |
| 12. Toelichtingen op antwoorden schrijven | 28.1 | 35.3 | 18.9 | 32.4 |
| 13. Opmerkingen in het opstel schrijven | 53.1 | 26.1 | 52.3 | 32.2 |

Waardering van de feedback. Er werden drie vragen gesteld die betrekking hadden op de perceptie en waardering van de feedback door gebruikers. De vraagstelling was in beide condities gelijk. De antwoordmogelijkheden varieerden echter. In de experimentele conditie hadden leerlingen vier antwoordmogelijkheden tot hun beschikking, waarbij het aantal commentatoren voor wie de vraag positief beantwoord kon worden varieerde van 3 tot 0. In de controleconditie kozen de leerlingen uit de antwoordmogelijkheden ja!!, ja!, nee! en nee!! In tabel 6.11. is samengevat hoeveel procent van de leerlingen niet tevreden was over de feedback die zij kregen. In de experimentele conditie achtten wij het onderwijs geslaagd als er tenminste over twee commentatoren tevredenheid bestond. De antwoordmogelijkheden in de controleconditie maakten het leggen van een cesuur niet moeilijk: 'nee!' en 'nee!!' duidden op ontevredenheid.

Tabel 6.11.: Ontevredenheid over de ontvangen feedback. Het percentage leerlingen dat het commentaar van twee of drie medeleerlingen negatief beoordeelde (E) en het percentage leerlingen dat de feedback van de docent negatief beoordeelde (C). (E=experimentele groep, C=controlegroep)

| | cursus 2 | | cursus 4 | |
|---|---|---|---|---|
| | E | C | E | C |
| Duidelijkheid | 11.6 | 22.6 | 10.9 | 5.3 |
| Zorgvuldigheid | 27.7 | 16.2 | 16.0 | 2.7 |
| Behulpzaamheid | 33.2 | 21.0 | 25.5 | 1.5 |

Uit tabel 6.11. is op te maken dat leerlingen in de controleconditie meer problemen hadden met de betekenis van het commentaar in de tweede cursus dan hun collega's in de experimentele conditie, maar desondanks de feedback op de twee andere aspecten positief beoordeelden. In de vierde cursus is de waardering in beide groepen in alle opzichten toegenomen. De leerlingen in de controleconditie hebben de feedback leren begrijpen en zijn zeer tevreden over de mate van behulpzaamheid. In de experimentele conditie is het oordeel over de zorgvuldigheid en de behulpzaamheid gunstiger.

Commentaar verwerken. Nadat de leerlingen het commentaar op hun opstel terugkregen, moesten zij dit samenvatten, beoordelen en verwerken in een herschrijfplan (tabel 6.12.).

Het blijkt dat de taken van de fase Commentaar Verwerken goed geïmplementeerd zijn. In de beide condities zijn in de vierde cursus wat meer leerlingen die zeggen het commentaar niet te hebben samengevat en niet te hebben becommentarieerd.

Tabel 6.12.: Deelname van leerlingen aan de taken uit de fase van Commentaar Verwerken. Het percentage leerlingen dat de activiteit <u>niet</u> verrichtte (E=experimentele groep, C=controlegroep)

| Taken | Cursus 2 E | Cursus 2 C | Cursus 4 E | Cursus 4 C |
|---|---|---|---|---|
| 1. Samenvatten van het commentaar | 10.7 | 12.0 | 5.8 | 12.4 |
| 2. Becommentariëren van het commentaar | 13.7 | 15.8 | 11.4 | 19.4 |
| 3. Opstellen van een herschrijfplan | 5.9 | 7.8 | 4.4 | 2.7 |
| 4. Schrijven van een definitieve versie | 3.6 | 5.4 | 2.8 | 1.9 |

In de handleiding voor de docent was aangegeven, dat leerlingen niet verplicht waren in de vierde ronde deze twee activiteiten afzonderlijk te verrichten: ze mochten als ze dat verkozen deze twee activiteiten in één keer op papier zetten. Aan commentaarverwerkers vroegen we of de commentatoren anoniem gebleven waren. Ook nu bleek dat we er niet in geslaagd waren het commentaar anoniem te laten geven. In de eerste cursus kende 36% van de schrijvers hun commentatoren, in de vierde cursus was dat percentage gestegen tot 63%.

Conclusies over het vragenlijstonderzoek. Uit de resultaten van het vragenlijstonderzoek maakten wij op dat de twee onderwijsprocedures zeker voor wat de kern betreft sterk geïmplementeerd zijn. Dat neemt niet weg dat er aanzienlijke verschillen tussen klassen zijn. We berekenden ook enkele hier niet gerapporteerde indices, waaruit bleek dat de mate van implementatie docentafhankelijk was. We berekenden een index voor aanbodverlies (het aantal activiteiten dat volgens 25% van de leerlingen van een klas niet is opgegeven) en een index voor deelnameverlies (het aantal lesactiviteiten waaraan een kwart van de leerlingen van een klas volgens eigen zeggen niet deelnam). Uit deze gegevens blijkt dat er tussen docenten en tussen klassen van docenten wat deze indices betreft soms grote verschillen zijn. Bij sommige docenten rapporteren relatief meer leerlingen uit beide klassen meer aanbod- en deelnameverlies. Bij een andere docent is het een van beide klassen waarin relatief veel aanbod- en deelnameverlies werd gerapporteerd. Sommige docenten hadden blijkbaar moeite met beide onderwijsprogramma's of hun beide klassen, andere docenten met een van beide klassen of een van beide onderwijsprogramma's.

Een belangrijk element van het experimentele onderwijsprogramma was het verrichten van de commentaartaak. Volgens de vragenlijstgegevens is deze sterk geïmplementeerd en zijn er steeds minder leerlingen (tabel 6.11.) die klachten hebben over de feedback die zij kregen: ook secundaire commentaartaken raakten beter in de werkwijze in klassen opgenomen. Leerlingen gaven aan veel tijd en energie in andermans opstellen te steken.

De kern van de feedbacktaak werd volgens feedbackgebruikers in beide condities uitgevoerd. Ze ontvingen van hun docent of van twee medeleerlingen de antwoorden op het reactie- en het commentaarformulier, en omschrijvingen van een eerste indruk en van het doel van de schrijver. Als we de feedbackgegevens uit tabel 6.10. rangordenen volgens het percentage leerlingen dat geen feedback op een taak zegt te hebben ontvangen, dan blijken de verschillen tussen de condities zich voor te doen na deze vier kerntaken. In de experimentele conditie ontvingen leerlingen voorts toelichtingen op de antwoorden van het commentaarformulier en reactietekens in de kantlijn van het opstel. In de controleconditie troffen leerlingen vooral nog verwijsnummers aan van commentaarvragen in het opstel en opmerkingen in het opstel. Behalve de kwantiteit van de feedbackgevers, zullen de twee condities tengevolge van het verschil in de persoon en kwaliteit van de feedbackgever ook hebben verschild in de soort en inhoud van secundaire feedback.

Er zijn verschillen tussen de condities wat de waardering van de feedback betreft (tabel 6.11.). Er zijn ook verschillen tussen de twee momenten. In beide condities neemt het aantal ontevreden leerlingen af. De docentfeedback is op beide momenten zorgvuldiger en biedt meer hulp dan leerlingenfeedback. Tijdens de tweede cursus waren leerlingen in de controle conditie minder over de duidelijkheid van de feedback van de docent te spreken.

In de vragenlijst over de vierde cursus namen we een vraag op over de mate waarin de herschreven versie van de eerste versie verschilde: 43% van de leerlingen in de experimentele conditie en 50% in de controleconditie zei dat de herschreven versie sterk van de eerste versie verschilde. Op de vraag of de commentaren veel invloed hadden gehad op de herschreven versie antwoordde 58% van de leerlingen in de experimentele groep bevestigend tegen 56% in de controlegroep. Veel invloed hebben op de herschreven versie wil blijkbaar nog niet zeggen dat die versie sterk verschilt van de eerste versie. Of de eerste versie in de controleconditie was van een hogere kwaliteit dan die in de experimentele conditie, òf in de experimentele conditie werden andere correcties aangebracht. In elk geval geldt voor beide condities dat de feedback die leerlingen kregen voor veel van hen ertoe deed.

## 6.5. Conclusies

In dit hoofdstuk werd beschreven welke procedures gehanteerd werden in de fase van dataverzamelen. Uit par. 6.2. bleek dat het hoofdbestanddeel van het onderzoek, het uitvoeren van de onderwijsprogramma's, in handen was van ervaren docenten Nederlands, werkzaam op een achttal scholen die varieerden in grootte en gezindheid. Zij voerden voorzover wij na konden gaan de programma's vrij nauwkeurig uit. Uit de door een

aantal van hen bijgehouden logboeken, bleek een duidelijk beeld te schetsen van de praktijk van de onderwijsprogramma's (par. 6.3.). Uit observaties en vragenlijstgegevens bleken beide onderwijsprogramma's goed geïmplementeerd. Uit de logboeken en de vragenlijstgegevens konden we ook enkele voorzichtige aanwijzingen verkrijgen over de verschillen tussen beide condities. Het bleek uit de logboeken dat er in de lessen waarin docenten de door hen becommentarieerde opstellen teruggaven in de controleconditie, veel onrust was. Leerlingen vroegen toelichtingen en nadere uitleg. Toch blijkt uit de antwoorden van leerlingen op vragen in de vragenlijst over hun oordelen over de feedback die zij kregen, dat het merendeel van de leerlingen daar heel tevreden over was. We vermoeden dat de interactie tussen docent en leerling een extra vorm van onderwijs is geweest die in de experimentele conditie niet plaatsvond. Uit de vragenlijstgegevens bleek ook dat de feedback van de docent waarschijnlijk andere signalen bevatte dan de feedback die leerlingen van elkaar kregen in de experimentele conditie. We denken dan aan de mate waarin door de docent gebruik is gemaakt van geschreven commentaar in de opstellen en van verwijsnummers: getallen in het opstel die corresponderen met vragen uit het commentaarformulier. Het is van belang dit soort gegevens over het functioneren van de onderwijsprogramma's te betrekken bij de interpretatie van de resultaten.

## 7. RESULTATEN

In dit hoofdstuk worden de resultaten van deze studie gepresenteerd. We brengen in herinnering dat er drie groepen afhankelijke variabelen zijn onderscheiden: Schrijfprestatievariabelen, Psychologische Variabelen en Schrijfprocesvariabelen. Het onderzoeksontwerp is een gerandomiseerd blokontwerp. Elf docenten (de blokken) hadden ieder twee klassen te onderwijzen. Per blok zijn de klassen via loting toegewezen aan een van beide condities. Voor de Psychologische Variabelen en de Schrijfprestatievariabelen beschikten we over de resultaten van drie meetmomenten: een voor-, tussen- en nameting. Voor de Schrijfprocesvariabelen werd als voor- en natoets een opstel hardop geschreven.

Dit hoofdstuk valt in zes delen uiteen. In par. 7.1. wordt ingegaan op de keuze van multivariate covariantie-analyse als statistische techniek om de hypothesen over de Schrijfprestatievariabelen en de Psychologische Variabelen te toetsen. In de daaropvolgende paragrafen worden de resultaten voor de hypothesen over de Schrijfprestatievariabelen (par. 7.2.), Psychologische Variabelen (par. 7.3.) en Schrijfprocesvariabelen (par. 7.4.) beschreven. In par. 7.5. komen enkele exploraties aan bod over de effecten van de factoren sexe en docent op de Psychologische Variabelen en de Schrijfprestatievariabelen, over de samenhang tussen deze afhankelijke variabelen en over de frequentie waarmee kenmerken van Doel- en Publiekgerichtheid in opstellen voorkomen. Het hoofdstuk wordt besloten met een concluderende samenvatting.

### 7.1. Multivariate covariantie-analyse

In par. 2.5. formuleerden we hypothesen over Psychologische Variabelen en over Schrijfprestaties. In plaats van elke afzonderlijke nulhypothese te toetsen, - bijvoorbeeld die voor elke opstelkwaliteit -, is het uit statistisch oogpunt verkieslijk eerst de multivariate nulhypothese te toetsen en pas bij verwerping daarvan elke afzonderlijke nulhypothese te onderzoeken. Men voorkomt daarmee dat het significantieniveau bij het toetsen van elke afzonderlijke hypothese ongemerkt groter wordt dan het per toets vastgestelde niveau (zie Mellenbergh, 1976).

Het gebruik van covariantie-analyse wordt aanbevolen in het geval dat men niet anders dan via statistische technieken kan corrigeren voor mogelijke initiële verschillen tussen de experimentele groepen. In onderwijsonderzoek is dat vaak het geval als gevolg van het ontbreken van de mogelijkheid de onderzoekseenheden via loting aan de condities toe te wijzen. In deze studie zou dat zo zijn als niet de klas maar de leerling de analyse-eenheid was geweest. Loting draagt ertoe bij dat eventueel beïnvloedende factoren als klassesfeer gelijkelijk verdeeld zijn in beide condities. Bovendien sluit

loting zelfselectie-effecten uit. Stel dat docenten zelf hadden kunnen bepalen welke van hun beide klassen de experimentele onderwijsprocedure zou volgen, op grond waarvan zouden zij hun keuze dan bepalen? Verwachtten zij in de ene klas beter met de experimentele leergang te kunnen werken dan in de andere klas?

De zekerheid waarmee men weet, dat beide condities qua rivaliserende achtergrondvariabelen gelijk zijn, neemt toe naarmate de condities meer via loting toegewezen onderzoekseenheden bevatten. In onze studie werd een groep gevormd door elf klassen. Dit is een vrij gering aantal, waardoor de kans op initiële verschillen die invloed zouden kunnen uitoefenen op de uiteindelijke resultaten van de natoetsen vrij groot is. Vandaar dat het toch wenselijk is voor deze initiële verschillen via de techniek van covariantie-analyse te corrigeren. Deze techniek neemt de correlatie tussen de afhankelijke variabele en een of meer covariaten in acht. De variantie in de natoetsscores die toe te schrijven is aan de covariaten wordt verwijderd uit de afhankelijke variabele voordat de significantietest wordt uitgevoerd.

In dit onderzoek werden de voortoetsscores opgevat als covariaten. De veronderstelling was dat de voortoetsscores zouden samenhangen met de natoetsscores: alleen dan is een regressielijn te schatten van de natoetsscores op de voortoetsscores, die de precisie van de variantie-analyse doet toenemen. Volgens Reichardt (in Cook en Campbell, 1979) is de vergroting van de precisie van covariantie boven gewone variantie-analyse gering als de correlatie tussen voor- en natoetsscores minder is dan .40. In tabel 7.1. worden de correlaties tussen voortoets en tussentoets en de voortoets en natoets gepresenteerd.

Uit de tabel blijkt dat covariantie-analyse op het analyseniveau van de klas zin heeft voor de Psychologische Variabelen en de Schrijfprestatievariabelen (schaalbeoordeeld). Hoewel bij Schrijfprestatievariabelen ook een aantal lage correlaties voorkomen in de kolom Totaal, zijn er voldoende correlaties van om en nabij .40 of hoger om covariantie-analyse te rechtvaardigen. Getoetst is met voor elke afhankelijke variabele de voortoetsscores op drie (Psychologische Variabelen) of vier (Schrijfprestatievariabelen) variabelen als voortoets, waarbij de met de afhankelijke variabele overeenstemmende variabele de eerste covariaat was.

Tabel 7.1.: Correlaties tussen voortoets-, tussentoets- en natoetsscores voor Psychologische Variabelen en Schrijfprestatievariabelen voor alle klassen (T) (n=22) en de klassen per conditie (E/C) (n=11), met tussen haakjes de correlaties voor alle leerlingen (T:n=264; E/C:n=132)

| Afhankelijke Variabelen | Voortoets-tussentoets | | | Voortoets-natoets | | |
|---|---|---|---|---|---|---|
| | T | E | C | T | E | C |
| **Psychologische Variabelen** | | | | | | |
| Angst-voor-het-niet-Kunnen-schrijven | .61 (.65) | .61 (.69) | .61 (.60) | .81 (.72) | .73 (.76) | .89 (.68) |
| Houding-jegens-het-Beoordeeld-worden | .53 (.59) | .33 (.54) | .72 (.66) | .46 (.60) | .30 (.50) | .72 (.69) |
| Houding-jegens het Schrijven | .58 (.72) | .14 (.73) | .72 (.71) | .49 (.72) | -.10 (.65) | .74 (.76) |
| **Schrijfprestatie-variabelen** | | | | | | |
| Doelgerichtheid | .38 (.20) | .41 (.18) | .41 (.23) | -.02 (.05) | .27 (.09) | -.35 (.09) |
| Publiekgerichtheid | -.02 (.14) | -.22 (.14) | .51 (.18) | .22 (.10) | .27 (.17) | .49 (-.07) |
| Opbouw | .36 (.25) | .19 (.27) | .70 (.24) | .52 (.09) | .62 (.06) | .37 (.18) |
| Stijl | .52 (.41) | .47 (.33) | .60 (.51) | .66 (.42) | .63 (.39) | .71 (.45) |

## 7.2. Resultaten Schrijfprestatievariabelen

In tabel 7.2. worden de gemiddelden en standaardafwijkingen gepresenteerd van de vier afhankelijke variabelen van het cluster Schrijfprestatievariabelen. Hoe hoger de score, hoe beter de prestatie. Het betreft tot klasseniveau geaggregeerde scores, verkregen via sommering van de individuele scores en deling van de somscores door het aantal elementen binnen de klas. De gerapporteerde standaardafwijkingen zijn indicatoren van de spreiding tussen klassescores, niet van de spreiding tussen leerlingscores.

Tabel 7.2.: Gemiddelden en standaardafwijkingen (tussen haakjes) van klassescores, binnen beide condities op de voortoets (1), tussentoets (2) en natoets (3) voor de afhankelijke variabelen Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl (alle vier via schaalbeoordeling)

| | Conditie | | | | | |
|---|---|---|---|---|---|---|
| | Experimenteel (n=11) | | | Controle (n=11) | | |
| | Toetsmoment | | | | | |
| Afhankelijke variabelen | 1 | 2 | 3 | 1 | 2 | 3 |
| Doelgerichtheid | 102.48 (4.23) | 105.42 (5.01) | 105.55 (5.97) | 100.92 (4.73) | 106.70 (5.89) | 106.72 (4.11) |
| Publiekgerichtheid | 103.37 (2.96) | 106.60 (5.79) | 100.30 (4.20) | 100.98 (2.40) | 108.52 (4.56) | 101.92 (3.08) |
| Opbouw | 91.54 (3.96) | 101.88 (4.50) | 106.23 (6.08) | 91.92 (2.47) | 99.53 (5.20) | 106.86 (5.84) |
| Stijl | 101.34 (3.98) | 104.85 (5.95) | 111.68 (3.69) | 101.87 (3.32) | 105.93 (4.49) | 112.17 (2.95) |

De covariantiematrices van beide groepen verschilden niet significant van elkaar (toets van Box). Gegevens over andere assumpties staan in tabel 7.3. Uit tabel 7.3. blijkt dat de niet lineaire associatiemaat eta ( $\eta$ ) in geen enkel geval significant van de lineaire correlatiecoëfficënt afwijkt. De relaties tussen de covariaten en de afhankelijke variabelen kunnen dus als lineair beschouwd worden. In de tweede helft van tabel 7.3. worden gegevens vermeld over de parallelliteit van de regressielijnen. Als de hellingshoek afhankelijk is van de conditie, zou dit een interactie tussen conditie en covariaat inhouden. De interactiecomponenten blijken echter in geen der gevallen statistisch significant. De verschillen tussen de gemiddelden van beide groepen zijn klein tot zeer klein. De multivariate nulhypothese kon dan ook voor noch de tussentoets-, noch de natoetsverschillen verworpen worden (MANCOVA: F=1.97, p=.16; F=.57, p=.69; df in beide gevallen 4/13). De toetsen zijn ook uitgevoerd met verwijdering van de twee klassen waarvan de docent zich vlak voor het einde van de derde schrijfronde terugtrok en die werden overgenomen door twee minder ervaren docenten (par. 6.3.). De uitslagen van deze toetsen doen echter niets af aan de conclusie (T2: F=1.96, df 4/11, p=.17; T3: F=.56, df 4/11, p=.71). De conclusie luidt dan ook dat er als gevolg van het al dan niet inschakelen van medeleerlingen geen verschillen in schrijfprestaties teweeg gebracht zijn.

Tabel 7.3.: Toetsingsgevens voor de assumptie van lineariteit en parallelliteit van de regressielijnen t.b.v. MANCOVA-toets voor de Schrijfprestatievariabelen DO (Doelgerichtheid), PU (Publiekgerichtheid), OP (Opbouw) en ST (Stijl)

| | | Lineairiteit | | | | | Parallelliteit | |
|---|---|---|---|---|---|---|---|---|
| Covariaat | Afhankelijke variabelen | η | r | F | vrijheidsgraden | p | Bijdrage van interactie tussen conditie en covariaat | p |
| DO voortoets | DO tussentoets | .6584 | .3914 | .6048 | 9 | .7704 | .00347 | .954 |
| | DO natoets | .4949 | .0244 | .3957 | 9 | .9127 | 1.78464 | .198 |
| PU voortoets | PU tussentoets | .7439 | -.0016 | 1.8592 | 8 | .1605 | 2.55343 | .127 |
| | PU natoets | .5344 | .2272 | .4913 | 8 | .4913 | .14577 | .707 |
| OP voortoets | OP tussentoets | .7666 | .3528 | 1.3738 | 9 | .3049 | 3.81065 | .067 |
| | OP natoets | .7376 | .5159 | .7481 | 9 | .7481 | .00834 | .928 |
| ST voortoets | ST tussentoets | .8370 | .5080 | 1.8069 | 9 | .1738 | .03123 | .862 |
| | ST natoets | .8516 | .6252 | 1.4877 | 9 | .2634 | .02066 | .887 |

Voor de variabelen Opbouw en Stijl is een gestage en grote verhoging in de scores te zien. Aan deze verschillen kan echter niet de conclusie verbonden worden dat beide onderwijsprocedures blijkbaar tot verhoging van de kwaliteiten van de opstellen leidden. Immers, in dit onderzoeksontwerp is niet voor groei gecontroleerd via een controlegroep zonder enige instructie. Zelfs de conclusie dat de leerlingen vooruit gegaan zijn, gaat te ver omdat de stelopdracht en derhalve de opstelschalen voor de voor- en tussentoets enerzijds en op de natoets anderzijds verschilden. In de discussie in hoofdstuk 8 komen we hierop terug.

## 7.3. Resultaten Psychologische Variabelen

In tabel 7.4. worden de gemiddelden en standaardafwijkingen gepresenteerd van de drie afhankelijke variabelen van de groep Psychologische Variabelen. Het betreft hier klassescores. Hoe hoger de score, hoe beter: minder angst en positievere houdingen.

De verschillen tussen de gemiddelden zijn wederom klein. Om deze verschillen statistisch te toetsen is gebruik gemaakt van een multivariate covariantie-analyse. Alvorens deze analyse te verrichten, is nagegaan of de gegevens aan twee belangrijke assumpties voldoen: parallelliteit en lineariteit van de regressielijnen binnen de condities.

Tabel 7.4.: Gemiddelden en standaardafwijkingen (tussen haken) van klassescores op de Psychologische Variabelen

| | Conditie | | | | | |
|---|---|---|---|---|---|---|
| | Experimenteel (n=11) | | | Controle (n=11) | | |
| | Toetsmoment | | | | | |
| Afhankelijke variabelen | 1 | 2 | 3 | 1 | 2 | 3 |
| Angst-voor-het-niet-Kunnen-schrijven | 47.20 (3.20) | 47.06 (4.12) | 47.69 (3.02) | 46.48 (3.03) | 46.72 (3.85) | 47.19 (3.54) |
| Houding-jegens-het-Beoordeeld-worden | 24.10 (2.54) | 25.17 (1.65) | 25.06 (3.45) | 24.53 (2.40) | 24.83 (2.42) | 24.81 (2.48) |
| Houding-jegens-Schrijven | 27.02 (1.52) | 24.79 (2.61) | 25.55 (4.01) | 25.30 (3.86) | 24.01 (3.47) | 24.57 (4.22) |

Tabel 7.5.: Toetsingsgegevens voor de assumptie van lineariteit en parallelliteit van regressielijnen t.b.v. MANCOVA-toets voor de Psychologische Variabelen: AK (Angst-voor-het-niet-Kunnen-schrijven), HB (Houding-jegens-het-Beoordeeld-worden), HS (Houding-jegens-Schrijven).

| | | Lineairiteit | | | | | Parallelliteit | |
|---|---|---|---|---|---|---|---|---|
| Covariaat | Afhankelijke variabelen | η | r | F | vrijheids-graden | p | Bijdrage van interactie tussen conditie en covariaat | p |
| AK voortoets | AK tussen-toets | .648 | .6006 | .1251 | 9 | .9979 | .00002 | .996 |
| | AK natoets | .859 | .8234 | .2762 | 9 | .9683 | 1.57690 | .225 |
| HB voortoets | HB tussen-toets | .750 | .5209 | 1.2392 | 7 | .3500 | 2.67419 | .119 |
| | HB natoets | .735 | .4661 | 1.3080 | 7 | .3204 | .45763 | .507 |
| HS voortoets | HS tussen-toetsen | .734 | .5927 | .6091 | 8 | .7548 | .45281 | .510 |
| | HS natoets | .687 | .4782 | .6894 | 8 | .6949 | 1.74830 | .203 |

Aan de assumpties van parallelliteit en lineairiteit van regressielijnen voldoen de gegevens (tabel 7.5.). Dat is niet het geval voor een derde assumptie, die van gelijke covariantiematrices in beide condities. Dit is nog wel zo bij de tussentoetsscores (toets van Box, F=1.10, df 21/1471, p=.34), maar niet bij de natoetsgegevens (F=1.79, df 21/1471, p=.015). Strikt genomen zou op deze gegevens geen multivariate cova-

riantie-analyse uitgevoerd mogen worden. We hebben dat toch gedaan: de F-waarde is namelijk klein maar wordt statistisch significant door het grote aantal vrijheidsgraden. De beide matrices verschillen dus niet sterk. De toetsingsresultaten wijzen uit dat de multivariate nulhypothese niet verworpen kan worden (tussentoets F=.12, p=.95; natoets F=.30, p=.83; df in beide gevallen 3/15). Ook hier voerden we de toetsen nog eens uit met verwijdering van de twee klassen die niet in ons ontwerp pasten, doordat een docenteffect geïntroduceerd werd toen een docent zijn twee klassen aan twee andere docenten overdroeg. De resultaten zijn echter vergelijkbaar (T2: F=.66, df 3/13, p=.59; T3: F=.19, df 3/13, p=.90). De conclusie over de Psychologische Variabelen moet hetzelfde luiden als die over de Schrijfprestatievariabelen: de twee onderwijsprocedures resulteren niet in verschillende houdingen in klassen.

## 7.4. Resultaten Schrijfprocesvariabelen

In par. 5.5. is beschreven op welke wijze de schrijfprocesgegevens van 22 leerlingen verkregen en bewerkt zijn. Na een inhoudelijke reductie van de 80 categorieën moest het aantal categorieën nogmaals teruggebracht worden in verband met het voorkomen van lege cellen. De tweede reductie leverde 22 (2x8-)tabellen op. Door de lage celfrequenties konden niet alle voorgenomen exploraties uitgevoerd worden (zie par. 2.5.). De 22 leerlingen bleken te weinig criteria te genereren om de categorie die hierop betrekking had te handhaven. Het onderscheid tussen publiekgerichte en neutrale opmerkingen moest eveneens vervallen om die reden: er werden nauwelijks publiekgerichte opmerkingen gemaakt. In tabel 7.6. staan de frequenties per conditie per categorie. De verwachting was dat leerlingen in de experimentele conditie op alle andere categorieën dan de restcategorie hoger zouden scoren dan de leerlingen in de controleconditie. Uit tabel 7.6. blijkt al dat dat niet het geval is.

De gegevens zijn geanalyseerd met behulp van loglineaire modelpassing (zie Vijn, 1982). Deze methode maakt het mogelijk de relaties tussen variabelen te onderzoeken door verschillende beschrijvingsmodellen tegen elkaar af te wegen. Tevens is het mogelijk, net als bij variantie-analyse, de bijdrage van de modeltermen te schatten.

Tabel 7.6.: Frequenties van Schrijfprocesvariabelen voor conditie en toetsmoment

| | Experimentele conditie | | Controle conditie | |
|---|---|---|---|---|
| Categorie | Voortoets | Natoets | Voortoets | Natoets |
| 1. Restcategorie | 2916 | 2808 | 3172 | 3027 |
| 2. Zelfinstructie | 192 | 222 | 217 | 163 |
| 3. Genereren | 1203 | 748 | 1184 | 836 |
| 4. Structureren | 128 | 194 | 136 | 187 |
| 5. Formuleringen kiezen | 44 | 30 | 33 | 50 |
| 6. Evalueren | 163 | 167 | 197 | 191 |
| 7. Veranderen van passages | 50 | 61 | 30 | 37 |
| 8. Veranderen van zinnen, woorden | 225 | 239 | 202 | 261 |

In ons model zijn er drie factoren: Conditie (1), Meetmoment (2) en Categorie (3). Gewoonlijk tracht men via successievelijk verwijderen van interactietermen tot een spaarzamer dan het verzadigde model te komen. In ons geval is echter de verwachting dat de interactie tussen conditie en meetmoment voor de categorieën 2 t/m 8 juist noodzakelijk is voor een adequate beschrijving van de data. In tabel 7.7. worden de gegevens van een aantal modellen gepresenteerd.

Tabel 7.7.: Aannemelijkheidsverhouding ($G^2$) met aantal vrijheidsgraden en overschrijdingskans voor een aantal modellen

| Model | $G^2$ | df | P |
|---|---|---|---|
| A (123) | | | |
| B (12) (13) (23) | 21.69 | 7 | .003 |
| C (12) (13) | 181.00 | 14 | .000 |
| D (12) (23) | 44.78 | 14 | .000 |
| E (13) (23) | 21.88 | 8 | .005 |

Uit de passingsgegevens in tabel 7.7. blijkt de interactieterm niet ongestraft uit het verzadigde model A verwijderd te kunnen worden. De interactieterm geeft aan dat er een verband is tussen Conditie, Tijdstip en Categorie: voor enkele of alle

categorieën is er een significant effect van de experimentele onderwijsprocedure. De interactieparameters zijn geschat, opdat nagegaan kan worden voor welke categorie er sprake is van een significante interactie. Deze parameterschattingen zijn getransformeerd tot standaardscores. Als de absolute waarde van zo'n standaardscore groter is dan 1.96, mag gesteld worden dat de desbetreffende cel bijdraagt tot de interactie (op 5%) (zie Vijn, 1982). In tabel 7.8. worden deze gestandaardiseerde parameterschattingen gepresenteerd.

Tabel 7.8.: Geschatte gestandaardiseerde interactieparameters onder model A, voor Conditie 1 (experimenteel) en Tijdstip 1 (voortoets) en Categorieën.

| | Categorie 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Conditie (1) x Cat. | -2.779 | .974 | -1.638 | -.472 | -.937 | -2.546 | 3.462 | -.220 |
| Tijdstip (1) x Cat. | 1.857 | 1.385 | 9.964 | -4.401 | .076 | .418 | -1.274 | -1.972 |
| Conditie x Tijdstip x Cat. | -1.021 | -3.532 | .678 | -1.067 | 2.517 | -.838 | -.204 | .967 |

De gegevens in tabel 7.8. laten zich als volgt lezen. Als de absolute waarde van de gestandaardiseerde parameterschatting in een cel groter is dan 1.96, wijkt hij significant ($p=.05$) af van 0 en draagt hij bij tot het effect op de afhankelijke variabele. Dit resulteert in de volgende interpretatie. Allereerst is het behoren tot een van beide condities er de oorzaak van dat de scores op categorie 1, 6 en 7 verschillen. Vervolgens is er het effect van tijd op de categorieën 3, 4 en 8, respectievelijk genereren, structureren en veranderen van woorden en zinnen. De leerlingen in beide groepen genereren minder en structureren en veranderen meer. Ten derde is er een effect van de experimentele onderwijsprocedure op categorie 2 en 5 (een interactie tussen tijdstip en conditie). De leerlingen in de experimentele conditie geven zichzelf op de natoets meer instructies dan op de voortoets, terwijl in de controleconditie het omgekeerde het geval is: de frequentie op de natoets is lager dan op de voortoets. Op categorie 5, overwegen van zinsbouw en woordkeus, is het patroon omgekeerd: hier blijken de leerlingen in de controleconditie van de onderwijsprocedure geprofiteerd te hebben.

Samengevat: de gemiddelde leerling genereert als gevolg van groei en/of het gemeenschappelijke in de onderwijsprogramma's minder items, maar structureert en verandert meer. De leerling in de experimentele conditie hanteert het schrijfproces als gevolg van het experimentele programma als een doelgericht proces, blijkens de toename van de frequentie van zelfinstructies. De leerling in de controleconditie houdt zich door het onderwijsprogramma meer met taalgebruikskeuzen bezig.

## 7.5. Psychologische Variabelen en Schrijfprestaties: nadere exploraties

In deze paragraaf beschouwen we enkele gegevens uit het onderzoek, die geen antwoord zijn op de primaire onderzoeksvragen, maar op zichzelf wel interessant zijn. We gaan in op drie thema's: analyses met de factor Sexe en Docent, een beschouwing van de gegevens van Doel- en Publiekgerichtheid, verkregen via analytische scoring, en enkele gegevens over correlationele verbanden tussen de afhankelijke variabelen.

In enkele onderzoeken die in het hoofdstuk over de onderzoeksliteratuur werden behandeld, bleek soms dat meisjes, soms dat jongens baat hadden bij feedback van leerlingen. Een verklaring voor deze interactie tussen Sexe en Programma werd in geen van deze onderzoeken gegeven. In één onderzoek was er sprake van een hoofdeffect voor sexe. Hoe het in ons onderzoek met de factor Sexe gesteld is, is nu voorwerp van aandacht.

Een eerste indicatie levert tabel 7.9., waarin de gemiddelden op de afhankelijke variabelen voor jongens en meisjes apart zijn vermeld. Een vergelijking levert enkele vermoedens op over hoofdeffecten van Sexe en interacties tussen Sexe en Programma. Zo lijken jongens minder Angst-voor-het-niet-Kunnen-schrijven te hebben dan meisjes en lijken meisjes meer plezier in schrijven te hebben dan jongens.

Tabel 7.9.: Gemiddelden en standaardafwijkingen voor jongens en meisjes apart op de Psychologische Variabelen en de Schrijfprestatievariabelen

| | Experimentele conditie | | | | | | Controle conditie | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | jongens[a] | | | meisjes[b] | | | jongens[c] | | | meisjes[d] | | |
| Afhankelijke variabele Toetsmoment | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Angst-voor-het-niet- | 49.39 | 49.88 | 49.98 | 45.47 | 45.01 | 45.95 | 46.49 | 48.46 | 48.06 | 46.09 | 45.32 | 48.24 |
| Kunnen-schrijven | 9.48 | 8.27 | 9.02 | 8.35 | 8.56 | 9.69 | 8.60 | 7.62 | 8.99 | 8.89 | 9.70 | 8.61 |
| Houding-jegens-Be- | 23.54 | 25.37 | 26.76 | 24.64 | 25.09 | 24.08 | 23.87 | 24.35 | 24.44 | 25.06 | 25.12 | 25.03 |
| oordeeld-worden | 6.29 | 5.78 | 4.60 | 5.67 | 4.84 | 6.02 | 5.73 | 5.68 | 6.30 | 5.45 | 5.46 | 5.71 |
| Houding-jegens- | 24.48 | 22.73 | 24.31 | 29.01 | 26.31 | 26.63 | 21.93 | 21.17 | 21.27 | 27.64 | 26.42 | 26.94 |
| Schrijven | 6.64 | 7.64 | 7.07 | 8.50 | 7.91 | 9.04 | 6.98 | 7.48 | 7.96 | 7.62 | 8.11 | 8.07 |
| Doelgerichtheid | 100.31 | 104.59 | 105.79 | 104.23 | 106.09 | 105.35 | 101.60 | 106.18 | 104.58 | 100.35 | 107.15 | 108.56 |
| (schaal) | 14.76 | 12.06 | 15.41 | 11.91 | 13.09 | 12.15 | 15.48 | 14.95 | 13.96 | 13.46 | 12.23 | 12.17 |
| Publiekgerichtheid | 102.83 | 106.55 | 100.19 | 103.77 | 106.64 | 100.39 | 99.57 | 106.76 | 100.23 | 102.20 | 110.04 | 103.38 |
| (schaal) | 10.01 | 11.76 | 8.21 | 8.18 | 12.47 | 9.0[illegible] | 8.61 | 11.19 | 7.87 | 8.49 | 9.27 | 7.27 |
| Opbouw (schaal) | 92.94 | 103.13 | 108.10 | 90.40 | 100.36 | 104.71 | 93.58 | 101.30 | 106.15 | 90.50 | 98.02 | 107.47 |
| | 9.86 | 10.34 | 11.76 | 9.42 | 9.85 | 12.04 | 10.59 | 12.67 | 12.81 | 7.81 | 9.76 | 10.61 |
| Stijl (schaal) | 100.99 | 105.37 | 112.49 | 101.62 | 104.42 | 111.03 | 101.52 | 103.68 | 110.21 | 102.40 | 107.86 | 113.86 |
| | 9.83 | 11.62 | 7.42 | 9.53 | 9.65 | 8.21 | 10.36 | 11.53 | 7.89 | 9.90 | 8.82 | 5.88 |

a) n = 54-59
b) n = 68-73
c) n = 56-61
d) n = 66-71

Wat de Schrijfprestatievariabelen betreft, lijkt er zich een interactie tussen Programma en Sexe voor te doen: meisjes in de controleconditie lijken meer van het onderwijs geprofiteerd te hebben dan meisjes in de experimentele conditie. Voor jongens is het beeld minder duidelijk. Voor sommige variabelen (Publiekgerichtheid bijvoorbeeld) lijkt de controleconditie, voor andere de experimentele conditie het gunstigst voor jongens.

Om deze vermoedens over de relatie tussen Programma en Sexe te toetsen, zijn multivariate covariantie-analyses uitgevoerd op een model waarin naast de factoren Programma en Sexe ook de factor Docent is opgenomen. De analyses betroffen zowel de tussentoets als de natoets. Als covariaten werden dezelfde variabelen gebruikt als in de analyses in par. 7.2. en 7.3. De analyses werden verricht op leerlingenniveau. Dit heeft tot gevolg, door de intraclasscorrelatie, dat het aantal vrijheidsgraden waarmee getoetst is te groot is. Uitgaande van de grootste intraclasscorrelatie van .07 (voor de kwaliteit Stijl) zou een reductie van ongeveer tweederde van het aantal vrijheidsgraden gepast zijn (Kish, 1965, p.162). Omdat het hier echter exploraties betreft, laten we die reductie achterwege. Alle toetsen zijn ook uitgevoerd voor een kleinere groep leerlingen, namelijk met uitzondering van de twee klassen op één school, die door twee docenten werden overgenomen. De resultaten verschilden niet met die van de analyses op het volledige bestand. De bevindingen van de laatste analyses worden hieronder gerapporteerd.

Uit alle vier de toetsen bleek een statistisch significant hoofdeffect van de factor Docent ($p \leq .05$). Dit wil zeggen dat er significante verschillen zijn op het moment van de tussentoets en de natoets tussen per docent 'gepaarde' klassen. Het wil niet zeggen dat de ene docent met zijn of haar beide klassen beter of slechter gepresteerd heeft dan de andere docent. Er kunnen ook schoolgebonden factoren zijn waardoor twee klassen op de ene school beter presteren dan twee klassen op een andere school. Uit een inspectie van de univariate analyses bleek dat het docenteffect zich voordoet op alle Schrijfprestatievariabelen (tussentoets en natoets).

Voor de Psychologische Variabelen treedt het docenteffect op voor Houding-jegens-Schrijven (T2: $F=2.69$, df 10/162, $p=.005$; T3: $F=3.17$, df 10/153, $p=.000$), Houding-jegens-Beoordeeld-worden (T3: $F=3.41$, df 10/153, $p=.000$) en Angst-voor-het-niet-Kunnen-schrijven (T3: $F=2.93$, df 10/153, $p=.000$).

Naast het hoofdeffect Docent bleek voor de Psychologische Variabelen een hoofdeffect voor Sexe op te treden ($p \leq .05$). Uit de univariate analyses bleek dit Angst-voor-het-niet-Kunnen-schrijven te betreffen (T2: $F=5.84$, df 1/162, $p=.017$) en Houding-jegens-Schrijven (T3: $F=5.60$, df 1/153, $p=.019$). Hoewel er dus enige ondersteuning is voor het vermoeden dat jongens zeggen minder angst te hebben en meisjes meer plezier in schrijven rapporteren, is het gebrek aan statistische significantie op de tussentoets (HS) en de natoets (AK) een teken dat een dergelijke uitspraak nog niet gedaan kan worden. Een sexe-effect werd ook aangetoond voor de Schrijfprestatie-

variabelen op de tussentoets (F=4.212, df 4/212, p=.01). Uit de univariate F-toetsen blijkt het de variabele Opbouw te betreffen (F=4.72, df 1/215, p=.03). Op de natoets is het sexe-effect in de multivariate toets niet meer significant. Alle gegevens met betrekking tot de factor Sexe overziend, moeten we besluiten dat er geen eenduidige verbanden tussen Sexe en de afhankelijke variabelen te constateren zijn.

Er bleek zich een interactie voor te doen tussen Docent en Programma (Psychologische Variabelen). Dit houdt in, dat de ene leraar succes had met de experimentele onderwijsprocedure voor een zekere afhankelijke variabele, een andere leraar juist met de controleprocedure. De interactie was op beide toetsmomenten significant, maar de absolute F-waarde was klein (T2: F=1.68, df 30/470, p=.014; T3: F=1.82, df 30/444, p=.006). Uit de univariate analyses bleek er voor de tussentoets geen afhankelijke variabele te zijn waarbij dit interactie-effect statistisch significant was op het 5%-niveau; voor de natoets betrof het de variabele Houding-jegens-Schrijven (F=2.65, df 10/153, p=.005). In één van de acht toetsen waarvan we enkele resultaten hiervoor rapporteerden, bleek een effect van de factor Programma. Op het tweede toetsmoment, de tussentoets, waren de verschillen tussen de condities voor Schrijfprestatievariabelen statistisch significant (F=3.47, df 4/212, p=.01). Uit de univariate analyses bleek het te gaan om de variabele Opbouw (F=4.34, df 1/215, p=.04). Noch de resultaten van de natoets (F=.76, df 4/212, p=.55), noch die van de analyse voor de tussentoets voor de gereduceerde groep (één school verwijderd) (F=2.18, df 4/192, p=.07) stemden hiermee overeen. Daarom houden wij ons bij onze conclusie uit par. 7.2., waar we stelden dat het al dan niet inschakelen van leerlingen geen verschillen in schrijfprestaties teweeg gebracht heeft.

Samenvattend kan gesteld worden dat er geen consistent patroon aangetoond kon worden van statistisch significante hoofdeffecten van Sexe, noch van interactie-effecten van Sexe en Programma. Geen van de onderwijsprogramma's bevoordeelt een van beide sexen. Wel bleek een hoofdeffect van Docent, dat wil zeggen dat sommige paren van klassen significant beter presteerden dan andere paren. Het effect treedt op voor vrijwel alle Schrijfprestatievariabelen, alsmede voor Psychologische Variabelen: Houding-jegens-Beoordeeld-worden en Houding-jegens-Schrijven. Voor de laatstgenoemde variabele werd ook een interactie tussen Docent en Programma aangetoond: de ene docent scoort met zijn experimentele klas hoger dan met zijn controleklas, bij een andere docent is het omgekeerd.

Een tweede exploratie betreft twee Schrijfprestatievariabelen: Doel- en Publiekgerichtheid. Welke kenmerken brachten leerlingen in hun teksten aan? Is er iets te zeggen over de verwervingsvolgorde? Zijn er mogelijk verschillen tussen condities? De frequenties van de onderscheiden kenmerken gunnen ons een blik achter de schermen. In de tabellen 7.10. en 7.11. wordt een overzicht gegeven van de frequenties waarmee de diverse kenmerken op de drie toetsmomenten voorkomen.

Tabel 7.10.: Kenmerken van publiekgerichtheid in inleidingen op drie toetsmomenten in percentages

| | Totaal (n=264) | | | Experimentele conditie (n=132) | | | Controleconditie (n=132) | | |
|---|---|---|---|---|---|---|---|---|---|
| Kenmerk / Toetsmoment | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 1. Het gebruik van je, jij en/of gebiedende wijs | 11 | 32 | 38 | 14 | 31 | 32 | 8 | 32 | 43 |
| 2. Het gebruik van ik | 51 | 43 | 53 | 56 | 47 | 49 | 46 | 39 | 58 |
| 3. Nadere kenmerken van de schrijver | 17 | 9 | 11 | 17 | 8 | 10 | 17 | 11 | 13 |
| 4. Overeenkomsten tussen schrijver en lezer | 6 | 12 | 22 | 5 | 8 | 19 | 8 | 15 | 24 |
| 5. Belang van informatie voor lezer onderstrepen | 27 | 35 | 35 | 30 | 33 | 39 | 23 | 37 | 31 |
| 6.* Emotief appel aan de lezer | 2 | 6 | 3 | 5 | 3 | 3 | 0 | 9 | 4 |
| 7. Een vraag | 23 | 53 | 51 | 21 | 50 | 49 | 26 | 56 | 52 |
| 8. Presentatie van onderwerp als iets ongewoons | 3 | 5 | 0 | .8 | 5 | .8 | 5 | 5 | 0 |
| 9. Beeldspraak | 1 | 2 | 1 | 0 | 2 | .8 | 2 | 2 | 2 |
| 10. Een uitroep | 4 | 8 | 16 | | | | | | |
| 11. Een anekdote | 21 | 11 | 25 | 24 | 10 | 21 | 17 | 12 | 30 |
| 12. Titel die inhoud en/of structuur aangeeft | 61 | 75 | 70 | 63 | 74 | 73 | 58 | 75 | 67 |
| 13. Identificatie van communicatief kader | 2 | 4 | 8 | 2 | 5 | 4 | 2 | 8 | 12 |
| 14. Identificatie van het tekstthema | 30 | 43 | 51 | 35 | 42 | 55 | 25 | 44 | 46 |
| 15. Informatie over de rest van het betoog | 4 | 12 | 7 | 2 | 14 | 6 | 7 | 11 | 8 |
| 16. Metacommunicatief karakter van de inleiding | 40 | 56 | 49 | 37 | 55 | 53 | 43 | 56 | 46 |
| 17. Leesbaarheid van de zinnen | 67 | 74 | 75 | 70 | 68 | 79 | 64 | 79 | 71 |
| 18. Begrijpelijkheid van de woorden | 86 | 81 | 91 | 89 | 86 | 90 | 83 | 75 | 92 |

* Niet betrouwbaar gescoord (par.5.2.)

Volgens Cohen (Cohen, 1977, p.183) is een verschil van 20 à 25%, overeenkomend met een middelmatige effectgrootte, met het blote oog waarneembaar (h=.50). Als we ervan uitgaan dat een docent een geoefend waarnemer is van opstelkwaliteiten en we stellen de effectgrootte op .40, dan zou deze docent de volgende verschillen tussen de toetsmomenten hebben opgemerkt. Hij zou op geen enkel kenmerk verschillen tussen zijn experimentele klas en zijn controleklas hebben opgemerkt. Wat Doelgerichtheid betreft zou hij uit de vergelijking van de opstellen van de voortoets en de natoets hebben opgemerkt, dat meer opstellen argumentatief van aard waren en meer opstellen een heldere lijn bevatten. Echter, dit laatste item is onbetrouwbaar gescoord. Hij zou gemerkt hebben dat in minder opstellen het standpunt in de inleiding geponeerd wordt en dat het slot vaker gebruikt wordt om het standpunt expliciet te vermelden. Wat Publiekgerichtheid betreft zou hij op het twee-

**Tabel 7.11.: Kenmerken van doelgerichtheid in betogen op drie toetsmomenten in percentages**

| | Totaal (n=264) | | | Experimentele conditie (n=132) | | | Controleconditie (n=132) | | |
|---|---|---|---|---|---|---|---|---|---|
| Toetsmoment / Kenmerk | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 1. Overwegend argumentatief karakter | 40 | 40 | 62 | 45 | 36 | 60 | 35 | 43 | 64 |
| 2. Formulering van een standpunt | 77 | 80 | 63 | 80 | 77 | 65 | 73 | 82 | 61 |
| 2a Standpunt in inleiding | 44[1] | 38 | 25 | 49 | 40 | 22 | 40 | 36 | 27 |
| 2b Standpunt in kern | 38[1] | 45 | 36 | 39 | 44 | 39 | 36 | 46 | 34 |
| 2c Standpunt in slot | 11[1] | 23 | 26 | 14 | 21 | 25 | 8 | 26 | 27 |
| 3.* Poging tot aannemelijkheidmaking | 84[1] | 87 | 91 | 83 | 86 | 91 | 85 | 88 | 90 |
| 3a* Eigen ervaringen ('ik') | 20[1] | 15 | 8 | 22 | 14 | 10 | 18 | 15 | 7 |
| 3b* Ervaringen van klas ('zij') | 30[1] | 25 | 9 | 27 | 24 | 9 | 33 | 27 | 8 |
| 3c* Ervaringen van lezers ('jij') | 14[1] | 20 | 29 | 16 | 23 | 24 | 11 | 17 | 33 |
| 3d* Causale keten, proces | 16[1] | 21 | 14 | 17 | 20 | 16 | 14 | 23 | 12 |
| 3e Oorzaak noemen | 6[1] | 11 | 8 | 9 | 8 | 13 | 4 | 13 | 4 |
| 4. Voldoende onderbouwing van de aannemelijkheidmaking | 27[2] | 29 | 27 | 26 | 30 | 29 | 28 | 27 | 26 |
| 5. Centraliteit van het betoog | 46 | 44 | 32 | 49 | 42 | 32 | 43 | 47 | 31 |
| 6.* Heldere lijn | 36 | 42 | 55 | 40 | 41 | 49 | 32 | 42 | 60 |

1) Van de betogen waarin een standpunt geformuleerd werd

2) Van de betogen die een standpunt en een aannemelijkheidmaking bevatten

* Niet betrouwbaar gescoord (par.5.2.)

de toetsmoment een hogere frequentie van het gebruik van de tweede persoon (kenmerk 1) en van het gebruik van de vragende vorm (kenmerk 7) waarnemen. Zou hij de opstellen van het derde toetsmoment met die van de voortoets vergelijken, dan zou hij buiten de eerder genoemde vooruitgang bovendien een hogere frequentie van kenmerk 4, - het scheppen van een band tussen lezer en schrijver via overeenkomsten - en van kenmerk 14, het identificeren van het tekstthema bemerken.

Tenslotte vragen we aandacht voor enkele correlaties tussen afhankelijke variabelen. In tabel 7.1. zagen we dat schrijfprestatiescores op leerlingenniveau onstabiel van karakter bleken te zijn. Leerlingen die voor één van de beoordeelde kenmerken tot de top van het klassement behoorden, scoorden in een volgende toets laag. Men zou verwachten dat wie er de eerste keer in slaagt een doelgerichte tekst te schrijven, er de tweede keer minstens even goed in zou slagen, zeker als het een exact gelijke opdracht betreft (tussentoets) of een opdracht van eenzelfde karakter (natoets).

Ook berekend zijn de correlaties tussen de diverse Schrijfprestatievariabelen binnen toetsmomenten (zie tabel 5.5.). Deze bleken niet tot matig met elkaar samen te hangen. De correlaties variëren van -.03 (Opbouw en Publiekgerichtheid op de voortoets) tot .58 (Opbouw en Doelgerichtheid op de natoets), met een gemiddelde van .35.

Tot slot werpen we een oog op de correlaties tussen Psychologische Variabelen en Schrijfprestatievariabelen (tabel 7.12.). Deze blijken nauwelijks met elkaar samen te hangen. Er doen zich wel enkele statistisch significante correlaties voor, maar deze zijn zo laag dat de sterkte van positieve attituden slechts maximaal 9% van de variantie van de variabele Stijl verklaart. Stijl is de enige Schrijfprestatievariabele die met de drie Psychologische Variabelen stelselmatig op de tussentoets en natoets samenhangt. 'Schrijven' in de items van het attitude-instrument betekent voor leerlingen wellicht 'formuleren', 'gedachten in woorden uitdrukken'.

Tabel 7.12.: Correlaties tussen Psychologische Variabelen en Schrijfprestatievariabelen op drie toetsmomenten

| | | Schrijfprestatievariabelen Schaalbeoordeling | | | |
|---|---|---|---|---|---|
| Toetsmoment | Psychologische variabelen | Doelgerichtheid | Publiekgerichtheid | Opbouw | Stijl |
| | Angst-voor-het-niet-Kunnen-schrijven | .09 | .13 | .09 | .14 |
| Voortoets | Houding-jegens-Beoordeeld-worden | -.05 | .04 | -.03 | .12 |
| | Houding-jegens-Schrijven | .08 | .20** | -.01 | .14 |
| | Angst-voor-het-niet-Kunnen-schrijven | .01 | .05 | .13 | .21** |
| Tussentoets | Houding-jegens-Beoordeeld-worden | -.09 | .15 | .09 | .16* |
| | Houding-jegens-Schrijven | -.03 | .09 | .11 | .21** |
| | Angst-voor-het-niet-Kunnen-schrijven | .14 | .15* | .17 | .29** |
| Natoets | Houding-jegens-Beoordeeld worden | .10 | .12 | .12 | .15* |
| | Houding-jegens-Schrijven | .09 | .17 | .16 | .30** |

* p=.01

** p=.001

## 7.6. Conclusies

In dit hoofdstuk zijn de resultaten gepresenteerd van het onderzoek naar effecten van leerlingenrespons op Psychologische Variabelen, Schrijfprestatievariabelen en Schrijfprocesvariabelen. Tevens zijn enkele exploraties verricht op de effecten van Sexe en Docent, de relatie tussen Psychologische Variabelen en Schrijfprestatievariabelen en de frequenties van kenmerken van Doel- en Publiekgerichtheid. De resultaten laten zich als volgt kort samenvatten.

De multivariate nulhypothesen over a) Psychologische variabelen en b) Schrijfprestatievariabelen konden niet verworpen worden. Klassen die een jaar lang via leerlingenrespons onderwezen werden scoorden niet statistisch significant anders dan klassen die docentfeedback kregen. Weliswaar was er voor een paar Schrijfprestatievariabelen, - Opbouw en Stijl - een forse vooruitgang in de scores, maar deze vooruitgang was in beide condities opgetreden. Of de vooruitgang aan de cursussen kan worden toegeschreven, is onzeker.

De data van de schrijfprocessen zijn beschreven met behulp van loglineaire modellen. Uit de passingsgegevens van deze modellen bleek alleen het verzadigde model te passen. De leerlingen bleken, ongeacht de conditie waarin ze les kregen, op de natoets minder te genereren en meer te structureren. De leerlingen in de experimentele conditie lijken het schrijfproces meer te sturen, blijkens de toename van de frequentie van zelfinstructies. De leerlingen in de controleconditie zijn zich meer op taalgebruikskeuzen gaan fixeren.

Uit de exploratie bleek dat jongens en meisjes niet systematisch van elkaar verschillen op Psychologische Variabelen en Schrijfprestatievariabelen. Er bleek ook geen differentieel effect van beide onderwijsprogramma's voor jongens en meisjes (interactie-effect tussen Sexe en Programma). Er is wel een effect van Docent geconstateerd: de prestaties en sommige attituden van twee klassen van eenzelfde docent zijn significant verschillend van die van twee andere klassen van een andere docent. Het ligt voor de hand dit verschil toe te schrijven aan de wijze waarop docenten variëren in hun instructiepraktijk. Het zouden ook schoolvariabelen kunnen zijn, die hier hun invloed uitoefenden. Er werd ook een interactie geconstateerd tussen Docent en Programma voor de variabele Houding-jegens-Schrijven.

Uit de exploraties bleek dat de Schrijfprestatievariabelen onderling niet of matig samenhingen (gemiddeld .35, variërend van -.03 tot .58). Eerder (tabel 7.1.) was al gebleken dat de Schrijfprestatievariabelen zeer onstabiel zijn. De correlaties tussen Schrijfprestatievariabelen en Psychologische Variabelen waren uitermate laag.

Een beschouwing van de resultaten van de scoring van Doelgerichtheid en Publiekgerichtheid op itemniveau, leverde het inzicht op dat er in het onderzoeksjaar voor een item van Doelgerichtheid (overwegend argumentatief karakter), vooruit-

gang geboekt werd. Voor Publiekgerichtheid was de oogst rijker. Van een viertal middelen om Publiekgerichtheid te bewerkstelligen werd geconcludeerd dat leerlingen deze zich aan het eigen maken waren tijdens het onderzoeksjaar.

Al met al zijn de resultaten verrassend eenduidig, maar onverwacht, gezien de hooggespannen verwachtingen van de practici in hoofdstuk 2 en de onderwijskundige verklaringen. Het stelonderwijsprogramma 'leerlingenrespons' heeft zich qua effecten op de schrijfprestaties en houdingen met betrekking tot stellen niet onderscheiden van het onderwijsprogramma 'docentfeedback'. Vanuit praktisch oogpunt moet en kàn een afweging gemaakt worden tussen de feedback van één docent die voor een opstel bijna een halfuur moet uittrekken of de feedback van drie medeleerlingen, die ieder een half uur à drie kwartier per opstel besteden. Het is de vraag hoe men deze resultaten dient te interpreteren: is de uitslag 0-0 (geen van beide programma's heeft effect) of 1-1 (beide programma's hebben goed gewerkt)? Een antwoord op deze vraag bewaren we voor het volgende (slot-)hoofdstuk.

# 8. DISCUSSIE

In de voorafgaande hoofdstukken deden we verslag van een veld-experiment, waarin twee stelonderwijsprogramma's met elkaar geconfronteerd werden. De programma's verschilden op één hoofdvariabele. In de experimentele conditie gaven leerlingen op elkaars opstellen commentaar (leerlingenrespons), in de controleconditie verrichtte de docent de feedbacktaak. Op basis van een analyse van publikaties over leerlingenrespons werd verwacht: 1) dat schrijfprestaties op de aspecten Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl zouden profiteren van de leerlingenrespons; 2) dat attituden jegens schrijven en beoordeeld worden positiever zouden worden en dat het vertrouwen in eigen schrijfvermogen zou toenemen als gevolg van leerlingenrespons; 3) dat schrijfprocessen zouden veranderen in de richting van meer doelgerichte probleemoplossingsprocessen, waarin kenmerken van de instructie een rol zouden spelen. De resultaten van het experiment rapporteerden we in hoofdstuk 7: voor Schrijfprestatievariabelen en Psychologische Variabelen troffen we geen verschillen aan tussen de condities, voor Schrijfprocesvariabelen wel.

In dit hoofdstuk bespreken we het onderzoek. Eerst (par. 8.1.) gaan we na in hoeverre de conclusies van het onderzoek valide zijn. Een aantal validiteitsbedreigende factoren worden besproken en hun gevolgen voor de houdbaarheid van onze conclusies. Daarna (par. 8.2.) gaan we in op de resultaten. We gaan na welke mogelijke verklaringen er zijn voor onze conclusies, en bespreken de houdbaarheid van die verklaringen en de eventuele consequenties. Na de bespreking der hoofdresultaten volgen de nevenresultaten (par. 8.3.): een aantal bevindingen die niet de essentie van ons toetsingsonderzoek vormden, doch die zó belangrijk zijn, dat ze hier besproken worden, in verband met hun importantie voor toekomstig didactisch onderzoek op het gebied van het onderwijs Nederlands. Tenslotte (par. 8.4.) bespreken we de consequenties voor de praktijk van het vak Nederlands. Wat kunnen we de practicus aanraden?

## 8.1. De validiteit van de conclusies

De uitslag van de statistische toetsing luidde (hoofdstuk 7) dat er geen significant verschil was tussen de scores in beide condities op Schrijfprestatievariabelen en Psychologische Variabelen. De verschillen in twee onderwijsprogramma's hingen kennelijk niet samen met verschillen in scores op de afhankelijke variabelen. Hoeveel waarde kunnen we hechten aan deze conclusie? Zijn er geen omstandigheden in ons onderzoek die deze conclusie op losse schroeven zetten? Cook en Campbell (1979) noemen zeven redenen waarom men zou kunnen twijfelen aan de waarde van de beslissing om van 'geen effect' te spreken. We gaan hieronder na of en in hoeverre wij op basis van een of meer van die factoren zouden moeten twijfelen aan de validiteit van de statistische beslissing.

Een eerste factor die Cook en Campbell noemen is een gering onderscheidingsvermogen van een statistische toets. In hoofdstuk 3 uitten wij kritiek op de daar gerapporteerde onderzoeken wegens het geringe onderscheidingsvermogen van de statistische testen die gebruikt werden, als gevolg van een betrekkelijk gering aantal waarnemingen. Onze onderzoeksopzet kan men een dergelijk verwijt ook maken. We analyseerden de resultaten immers op klasseniveau: als we een middelmatig effect verwachtten, zouden we 20% kans hebben een aanwezig effect vast te stellen. Een dergelijk onderzoek had men achterwege moeten laten, zouden we in hoofdstuk 3 hebben gezegd. Echter, we beschikten op een ander analyseniveau wel over voldoende waarnemingen: elke klassescore was opgebouwd uit twaalf leerlingscores. Op dit niveau zouden we een onderscheidingsvermogen van 90% bereiken. De verschillen tussen de scores in beide condities waren echter gering. Bij een toetsing van een model met Conditie, Docent en Sexe als factoren met een voor intraclasscorrelaties ongecorrigeerd aantal vrijheidsgraden kon geen systematisch statistisch significant verschil aangetoond worden voor Conditie.

Een tweede factor waardoor men zou kunnen twijfelen aan de validiteit, is de mate waarin voldaan is aan de assumpties van de gebruikte statistische toets, in ons geval multivariate covariantie-analyse. Wij menen daaraan ruime aandacht besteed te hebben. In hoofdstuk 7 lieten we zien dat we belangrijke assumpties controleerden, zoals die van parallelle en lineaire regressielijnen en gelijke covariantie-matrices.

Een derde factor, de kwestie van 'error-rate', behandelden we in 7.1., waarin we toelichtten waarom we een multivariate techniek toepasten in plaats van een univariate. Op deze wijze omzeilden we het gevaar van een ongemerkt toenemen van het significantieniveau als er meer afhankelijke variabelen zijn waarvoor telkens apart een toets uitgevoerd wordt.

Een vierde factor die een ongunstige invloed uitoefent op de houdbaarheid van de statistische conclusie, is de betrouwbaarheid van de metingen. Op het analyseniveau van de leerling bleken de scores op de Psychologische Variabelen wèl, maar die op de Schrijfprestatievariabelen nîet stabiel. De onbetrouwbaarheid is niet toe te schrijven aan de beoordelingen van de opstellen; deze zijn betrouwbaar tot stand gekomen (par. 5.1.). De prestatie van een leerling fluctueerde blijkbaar, zelfs al waren de toetsen die op twee tijdstippen werden afgenomen identiek. Doordat wij echter de scores aggregeerden tot op klasseniveau, verkregen we stabielere scores, hetgeen af te lezen is uit tabel 7.1. Voor de analyses op klasseniveau heeft de onbetrouwbaarheid van de meting van Schrijfprestatievariabelen dus niet het nadelige gevolg van een toename van de foutenvariantie. Deze 'winst' gaat echter gepaard met een verlies aan vrijheidsgraden en onderscheidingsvermogen. De onbetrouwbaarheid van de metingen speelt op die wijze wel een rol. Bij de afweging van de ernst daarvan moeten de verschil-

len in de scores betrokken worden. Deze zijn zo klein dat wij de conclusie blijven onderschrijven.

Een vijfde factor is de mate van variatie in de implementatie van de onderwijsprogramma's in de experimentele en de controleconditie. In ons onderzoek trachtten we de implementatie zo goed mogelijk te standaardiseren door middel van een strak lessenplan (in de handleiding voor de docenten), zeer voorgestructureerd lesmateriaal, dat bovendien in grote mate zelfinstruerend was, en door middel van een terugkerend stramien van leeractiviteiten. Op verschillende wijzen is gecontroleerd of het lessenplan gevolgd werd en in welke mate leerlingen aan de programma's deelnamen. Uit gegevens hierover (par. 6.4.) bleek dat beide programma's zoals gepland werden uitgevoerd en dat in beide condities leerlingen in vrij sterke mate aan de programma's deelnamen. Wel bleken er verschillen tussen docenten. Van de ene docent rapporteerden beide klassen minder deelname aan taken dan de beide klassen van een andere docent. Variabiliteit tussen klassen binnen condities bleek onvermijdelijk als een onderwijsprogramma zo lang duurt, hoezeer we ook trachtten de programma's docentonafhankelijk te maken. Toch achten we de variabiliteit niet van dien aard, dat deze in ernstige mate zou bijdragen aan de foutenvariantie, waardoor de kans op het vaststellen van ware verschillen zou afnemen. De implementatiegraad was immers zo hoog, dat afwijkingen daarvan niet heel ernstig kunnen zijn.

Als zesde factor noemen Cook en Campbell mogelijke irrelevante eigenschappen van de onderzoekssituatie, die de scores op de afhankelijke variabelen beïnvloeden en de foutenvariantie doen toenemen. Een bron van storing kan de persoon van de docent zijn. Via de onderzoeksopzet (blokken op docent) boden we dit gevaar het hoofd. Een andere bron zou het overige onderwijs in het vak Nederlands kunnen zijn en het eventuele stelonderwijs bij andere vakken. Voor wat het vak Nederlands betreft hebben we getracht rivaliserende lesactiviteiten op het spoor te komen via docentlogboeken. Probleem was echter welke lesactiviteiten als storend gedefinieerd kunnen worden, omdat er nog zo weinig bekend is over factoren die schrijfprestaties beïnvloeden. Wat in andere schoolvakken aan stelonderricht gegeven is, is ons niet bekend. Helemaal uitsluiten kunnen we een dergelijke storingsbron dus niet. Wellicht zijn de scores op Psychologische Variabelen en Schrijfprestatievariabelen beïnvloed door irrelevante kenmerken van de onderzoeksomgeving. We schatten echter dat door de duur en intensiteit van de beide onderwijsprogramma's de rol van andere factoren niet groot geweest zal zijn.

De laatste factor die Cook en Campbell noemen is aselecte heterogeniteit van proefpersonen. Dit wil zeggen dat de proefpersonen in de condities verschillen op factoren die samenhangen met de afhankelijke variabelen. Als sexe bijvoorbeeld invloed zou hebben op de scores op de afhankelijke variabelen en de proportie jongens en meisjes zou in beide condities verschillen, terwijl deze factor niet vastgesteld

zou zijn in het onderzoek, dan zou de foutenvariantie hierdoor vergroot zijn. Op twee wijzen hebben we dit probleem het hoofd geboden. Ten eerste was onze onderzoekspopulatie vrij homogeen (derde klas van het H.A.V.O. en het V.W.O.). Ten tweede namen we die variabelen als covariaten op waarvan verwacht kon worden dat zij aan de afhankelijke variabelen gerelateerd zouden kunnen zijn: Schrijfprestatievariabelen en Psychologische Variabelen.

Alles overziend, zien we weinig redenen om sterk aan de conclusie te twijfelen: de experimentele variabele (feedback) covarieert niet met verschillen op Schrijfprestatievariabelen en Psychologische Variabelen. In het vervolg van dit hoofdstuk zullen we dan ook niet meer aan deze conclusie tornen.

Een enkel woord nog over het onderzoek naar effecten op Schrijfprocesvariabelen. De leerlingen die zich aanmeldden, - vrijwillig, op uitnodiging of na aandrang van hun docent - kunnen wegens deze zelfselectie niet representatief geacht worden voor de gehele groep leerlingen die aan het onderzoek deelnam. We weten immers niet op welke kenmerken de zelfgeselecteerde leerlingen verschillen van hun klasgenoten. Voor de vergelijking tussen de processen in beide condities heeft die zelfselectie geen gevolgen: leerlingen wisten bij aanmelding niet welk onderwijsprogramma zij zouden volgen en er is dus weinig reden om aan te nemen dat de leerlingen in de experimentele conditie anders zouden zijn dan die in de controleconditie. Wel moet voorzichtigheid betracht worden bij het verbinden van de resultaten van het Schrijfprocesonderzoek en het Schrijfprestatie-onderzoek als gevolg van zelfselectie in het Schrijfprocesonderzoek.

## 8.2. De resultaten verklaard: een poging

In deze paragraaf trachten we de drie hoofdconclusies nader te verklaren, namelijk dat verschillen in onderwijsprogramma's leidden tot verschillen in schrijfprocessen en niet tot verschillen in schrijfprestaties en attituden. We besteden eerst aandacht aan de veranderingen en verschillen in schrijfprocessen. Vervolgens staan we stil bij de resultaten op de Psychologische Variabelen. Het uitvoerigst behandelen we mogelijke relaties tussen kenmerken van de onderwijsprogramma's en de resultaten op de Schrijfprestatievariabelen.

De veranderingen in de schrijfprocessen in beide condities, - meer structureren en meer veranderen -, zijn uit oogpunt van de schrijfprocestheorie gunstig te noemen. Beide veranderingen geven blijk van 'reprocessing' (Scardamalia en Bereiter, 1986): ideeën en tekst worden bewerkt. Leerlingen genereren niet slechts associatief wat op een of andere wijze in het lange termijn geheugen opgeslagen ligt, maar vormen de gegenereerde items om, evalueren ze, ordenen ze. Een dergelijk proces zou de coherentie van de tekst ten goede moeten komen. Het frequenter optreden van 'reprocessing'-subgroepen, - denk-

processen die werken op reeds in andere processen gegenereerde elementen -, duidt erop dat het schrijfproces minder lineair en meer recursief verloopt. De leerlingen in de experimentele conditie maakten bovendien meer zelfinstruerende opmerkingen op de natoets, een mogelijke aanwijzing dat zij meer macht over het proces hebben, omdat ze het opdelen in subprocessen. Waarom leerlingen in de controleconditie echter vergeleken met de voortoets minder van dergelijke opmerkingen zijn gaan maken, is minder makkelijk te interpreteren. Wellicht moeten we die teruggang in verband brengen met de vooruitgang van het aantal formuleringskeuzen: leerlingen in de controleconditie genereren formuleringen, die zij vaker dan voorheen beoordelen voordat ze die aan het papier toevertrouwen. Als we beide bewegingen met elkaar in verband brengen, zouden leerlingen in de experimentele conditie meer procesgeoriënteerd werken en leerlingen in de controleconditie meer tekstgeoriënteerd. Het zijn echter kleine aantallen proefpersonen waarover het hier gaat, en de overige resultaten wijzen niet in deze richting. Analyse van de overige, wel verzamelde maar niet gecodeerde protocollen zou een steviger basis bieden voor een dergelijke interpretatie.

Waarom had leerlingenrespons geen invloed op scores op Psychologische Variabelen? De verwachting was dat leerlingen meer vertrouwen in hun eigen schrijfvermogen zouden krijgen, minder angst zouden hebben voor het beoordeeld worden en schrijven leuker zouden gaan vinden.

De positievere Houding-jegens-het-Beoordeeld-worden werd door practici (hoofdstuk 2) toegeschreven aan de veiliger taaksituatie bij leerlingenrespons, doordat de beoordelende docent uitgerangeerd is. De taaksituatie was echter door het toevoegen van een herschrijffase aan de instructiereeks in beide condities minder bedreigend dan in schrijfonderwijs gebruikelijk is. Daar is het schrijven van een opstel verbonden met een voor het rapport meetellende beoordeling. De scores op Houding-jegens-het-Beoordeeld-worden (tabel 7.4.) laten in beide condities een lichte stijging zien, maar niet zodanig dat men kan zeggen dat wellicht het toevoegen van een herschrijffase tot minder beoordelingsangst zal leiden, of het moet zijn dat gedurende het derde leerjaar van het H.A.V.O. en V.W.O. gewoonlijk de beoordelingsangst toeneemt en dat onze onderwijsprogramma's die toename afremden. We weten echter niets over het 'normale' niveau van Houding-jegens-het-Beoordeeld-worden. Leerlingenrespons leidt in elk geval niet tot vooruitgang op deze variabele.

Door het vergelijken van eigen met andermans produkten zouden leerlingen meer kennis over de eigen bekwaamheid verwerven, die zou leiden tot meer vertrouwen in eigen kunnen, ofwel lagere scores op de variabele Angst-voor-het-niet-Kunnen-schrijven. De scores (tabel 7.4.) vertoonden geen ontwikkeling over de tijd heen en geen verschillen tussen beide condities. Leerlingenrespons leidt blijkbaar niet tot groei in het vertrouwen in het eigen schrijfvermogen.

Doordat leerlingen goede teksten lezen en betere teksten gaan schrijven zouden leerlingen meer plezier in schrijven krijgen. Uit tabel 7.4. blijkt dat het plezier in schrijven in beide condities afnam en dat er nauwelijks verschillen tussen de condities waren. Bij het onderzoek betrokken docenten meldden ons tijdens het onderzoek dat leerlingen de cursus wel erg intensief begonnen te vinden. Wellicht kan daaraan de afname van het plezier in schrijven toegeschreven worden.

Dat leerlingenrespons niet leidde tot verschillen tussen de condities op de drie Psychologische Variabelen, kan toegeschreven worden aan de verschillen tussen de onderwijsprogramma's, die misschien niet groot genoeg waren. Op die redenering komen we terug als we de resultaten op de Schrijfprestatievariabelen behandelen. Voor de Psychologische Variabelen voegen we er nog twee mogelijke verklaringen aan toe. Ten eerste: uit de correlaties tussen leerlingenscores op verschillende tijdstippen blijkt dat de drie variabelen ieder een over een tijdsduur van een cursusjaar vrij stabiele trek representeren, die zich blijkbaar op deze leeftijd niet makkelijk laat beïnvloeden door een schrijfvaardigheidscursus. Ten tweede: de correlaties tussen Psychologische Variabelen en Schrijfprestatievariabelen (tabel 7.12.) zijn zeer laag en vaak niet eens statistisch significant. Als Psychologische Variabelen zouden veranderen, zoals bij Houding-jegens-Schrijven verondersteld werd, door de toegenomen kwaliteit van de eigen opstellen, is zo'n causaal verband door het ontbreken van covariantie al onmogelijk. Op de lage correlaties tussen Psychologische Variabelen en Schrijfprestatievariabelen komen we in par. 8.3. nog terug. Hier benadrukken wij dat er aanwijzingen zijn dat de Psychologische Variabelen stabiele trekken representeren en dat er geen verband is tussen deze variabelen en Schrijfprestatievariabelen. Deze twee uitspraken kunnen afzonderlijk of gezamenlijk verklaren waarom leerlingenrespons geen invloed heeft gehad op de scores op de drie Psychologische Variabelen.

Hoe is het ontbreken van verschillen op de Schrijfprestatievariabelen inhoudelijk te verklaren? Als eerste verklaring zou aangevoerd kunnen worden dat de onderwijsprogramma's bijzonder weinig verschilden. We hebben het ons niet gemakkelijk gemaakt door leerlingenrespons af te zetten tegen een 'gemiddelde' steldidactiek, maar construeerden een concurrent die veel meer is dan wat er gemiddeld aan stellen wordt gedaan. Hoewel de programma's zoveel mogelijk op elkaar leken, waren de verschillen tussen de twee programma's aanzienlijk. In de experimentele conditie lazen leerlingen elkaars opstellen, gaven er commentaar op, kregen van drie leerlingen feedback; in de controleconditie kregen de leerlingen alleen van hun docent commentaar. De leerlingen in de experimentele conditie besteedden meer tijd - twee lesuren per schrijfronde en driekwartier extra huiswerk - aan het onderwijsprogramma. Bovendien zijn er aanwijzingen uit de logboekanalyse en het vragenlijstonderzoek dat er zich nog meer verschillen voordeden tussen beide condities. Ten eerste bleken in de controle-

conditie de lessen waarin de docent de opstellen met feedback teruggaf, onrustig. Leerlingen vroegen voortdurend toelichtingen en de docent liep de klas rond om nadere informatie te verstrekken. Het is zeer waarschijnlijk dat in dit lesuur een vorm van extra onderwijs gegeven werd die de leerlingen in de experimentele conditie ontbeerden. Ten tweede bleek dat docenten meer dan leerlingen extra commentaar in de opstellen schreven en vaker vraagnummers van het feedbackinstrument in de kantlijn van het opstel plaatsten. De aard van de feedback kan dus verschillend zijn geweest voor beide condities. Hoe we deze twee verschillen moeten waarderen, blijft onzeker. Het extra onderwijs dat sommige leerlingen in de controleconditie kregen, geldt niet voor de hele klas. Die leerlingen die toelichtingen ontvingen zullen misschien bevoordeeld zijn, maar het verschil kan nooit tot een systematisch verschil verheven worden. Het effect van de extra geschreven toelichtingen door docenten, kan leerlingen geholpen hebben de accenten in de feedback te begrijpen, terwijl daarentegen leerlingen in de experimentele conditie zelf de weg moesten vinden in de grote hoeveelheid feedback van drie commentatoren. We achten de verschillen inherent aan de twee condities. Ze worden niet opgevat als de geplande verschillen verdoezelend. De verschillen tussen de condities zijn ons inziens groot genoeg om verschillen in schrijfprestaties te verwachten.

Een tweede verklaring voor het uitblijven van verschillen kan zijn dat in de experimentele conditie geleerd werd van het kritiek géven, maar dat de leerwinst ten opzichte van leerlingen uit de controleconditie teniet werd gedaan doordat de kwaliteit van de feedback in beide condities verschilde. Het is aannemelijk dat leerlingen in de controleconditie kwalitatief beter commentaar kregen, namelijk afkomstig van een deskundige, die veel tijd in een opstel stak, dan de leerlingen in de experimentele conditie. Toch bleek uit het vragenlijstonderzoek dat leerlingen in beide condities tevreden waren over de duidelijkheid, zorgvuldigheid en behulpzaamheid van de feedback. Nu kan tevredenheid wel opgevat worden als operationalisatie van de gepercipieerde kwaliteit van de feedback, we weten niet welke relatie er is tussen de waargenomen kwaliteit en de objectieve kwaliteiten, zoals bijvoorbeeld de verwerkbaarheid en de juistheid van het commentaar. Toch beschikken we hiermee over een aanwijzing dat er volgens het oordeel van leerlingen niet veel verschil is tussen de twee typen feedback die in ons onderzoek centraal stonden. Ook op andere vragen die betrekking hadden op het verwerken van de feedback en de invloed ervan op de herschreven versies, ontliepen de antwoorden in beide condities elkaar niet veel. We houden het erop dat de condities niet verschilden in de kwaliteit van de feedback die zij kregen. Dat betekent overigens ook, dat er twijfel is over de meerwaarde van 'peer feedback' die in de theorie (hoofdstuk 2) benadrukt wordt. Een nader onderzoek naar de kwaliteit en de effectiviteit van de

feedback die leerlingen en die docenten geven zou, in de vergelijking tussen leerlingenrespons en docentfeedback betrokken moeten worden; ons onderzoek bood daar gezien de praktische omstandigheden geen gelegenheid toe.

Het is zelfs de vraag of feedback er überhaupt toe doet. Weliswaar is feedback nuttig bij het reviseren van een bestaande, reeds geschreven tekst, maar bij het produceren van een nieuwe tekst is de oude feedback misschien geen adequate steun. Het schrijven van een nieuwe tekst is een nieuw probleemoplossingsproces, waarbij specifieke feedback bij een vorig oplossingsproces natuurlijk slechts zeer gedeeltelijk van nut is. We hebben één gegeven dat onze veronderstelling over de geringe waarde van feedback, ook als die van medeleerlingen afkomstig is, ondersteunt. We rangordenden de klassen binnen een conditie naar de proportie leerlingen die zich ontevreden toonden over de bruikbaarheid van de feedback en deden dat voor zowel de tweede als de vierde schrijfronde (vragenlijstonderzoek, par. 6.4.). We rangordenden de klassen tevens op de schrijfprestaties. Zowel voor de tussentoets als de natoets, voor alle Schrijfprestatievariabelen apart (Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl) en voor hun som. Vervolgens berekenden we de rangordecorrelaties tussen de mate waarin leerlingen klaagden over de feedback en de verschillende schrijfprestatierangordeningen. Van de tien rangordecorrelaties tussen waargenomen kwaliteit van de feedback en schrijfprestaties in de controleconditie was er slechts één significant. Kennelijk hangen verschillen tussen klassen met betrekking tot de waardering van de feedback niet samen met verschillen in de schrijfprestaties. In de experimentele conditie troffen we drie significante rangcorrelaties aan tussen de waargenomen bruikbaarheid en schrijfprestaties, nl. Stijl (2x) en totale schrijfprestatie. Deze gegevens, tezamen met die uit andere onderzoeken, waaruit blijkt dat de intensiteit, de toon, de wijze van presenteren, ja waar zelfs de aanwezigheid van feedback er niet toe zou doen (Wesdorp, 1983), doen ons veronderstellen dat feedback in schrijfonderwijs weleens een veel minder belangrijk instructiemiddel zou kunnen zijn, dan wij voorheen dachten. Ook 'peer feedback' zou weleens minder instructief kunnen zijn dan geopperd wordt door de practici, wier ideeën wij gebruikten voor onze theorie in hoofdstuk 2.

Wanneer het zo is, dat de feedbackverschillen tussen de condities geen prestatieverschillen bewerkstelligen, en wanneer het wellicht zo is dat feedback zelfs niets of nauwelijks bijdraagt tot prestatieverbetering, dan blijft er nog een verschil tussen de condities over, tengevolge waarvan prestatieverschillen verwacht werden. In de experimentele conditie gaven leerlingen wel, in de controleconditie geen commentaar. In de experimentele conditie hadden leerlingen dus meer kans om de criteria voor een goede tekst, tevens de doelstellingen van de cursus, te internaliseren. Men verwachtte veel van deze leeractiviteit. Uit secundaire gegevens

bleek ook wel dat leerlingen in de experimentele conditie meer vat hadden op de criteria dan hun collega's in de controleconditie. Vooral in de al eerder aangehaalde lessen waarin docenten de opstellen met hun commentaar teruggaven in de controleconditie, bleek dat de leerlingen veel toelichtingen behoefden. Dit zou erop kunnen duiden dat er in de experimentele conditie wel, in de controleconditie minder inzicht in de inhoud van de criteria was. Echter, in de experimentele conditie was het een spelregel, dat leerlingen niet met elkaar overlegden over de feedback, hoewel er dedactisch gezien veel voor te zeggen is, omdat in één-op-één interacties veel verduidelijkt kan worden. Het is dus best mogelijk dat ook zij niet zoveel inzicht hadden als hiervoor werd aangenomen. Anderzijds vonden leerlingen in beide condities de feedback duidelijk, behulpzaam en zorgvuldig. Daaruit leiden wij af dat leerlingen in beide condities toch criteria voor goede teksten verwierven. Wellicht leidde het geven van commentaar tot een voorsprong van de leerlingen in de experimentele conditie op de leerlingen in de controleconditie, maar is deze voorsprong later teniet gedaan. Doordat ze feedback kregen via deze criteria en doordat de instructieteksten waarin de criteria in samenhang gepresenteerd werden als achtergrondkennis gingen functioneren, zouden de leerlingen in de controleconditie het kennisniveau van de leerlingen in de experimentele conditie bereikt kunnen hebben. Maar evenals aan de waarde van feedback kan men twijfelen aan de waarde van kennis over criteria voor goede teksten voor het produceren van tekst. Weten waaraan een goede tekst moet voldoen, betekent nog niet dat men een goede tekst kan schrijven. Immers, de verwachtingen dat in de schrijfprocessen van de experimentele groep meer criteria gegenereerd zouden worden, kwamen niet uit. Leerlingen genereerden zo weinig criteria dat de betreffende cellen vrijwel leeg bleven. Wellicht is de veronderstelde relatie tussen kennis van criteria en schrijfprestaties helemaal niet zo sterk als pleitbezorgers van leerlingenrespons in hoofdstuk 2 schetsten. Deze bevinding sluit aan bij wat Rubin (1983) opmerkte: leerlingen weten wel wat er aan een tekst mankeert, maar kunnen de gebreken niet verhelpen. In een nader onderzoek zou de relatie tussen kennis van criteria en schrijfprestaties onderzocht moeten worden.

Samengevat komt het erop neer dat we ondanks de verschillen in de onderwijsprogramma's denken dat beide tot kennis over criteria leidden, zowel via de feedback over opstellen als via het vervullen van de feedbacktaak. Daardoor zou de overeenkomst tussen beide condities qua schrijfprestaties verklaard kunnen worden. Verder twijfelden we aan het nut van feedback voor de nieuwe schrijftaken en, daarmee geïmpliceerd, aan het nut van kennis over criteria voor het produceren van teksten. Gesteld dat we het met deze laatste stelling bij het rechte eind hebben, hoe is dan te verklaren dat de schrijfprestaties over het algemeen beter werden, zelfs zo, dat men moeilijk kan volhouden dat de vooruitgang uitsluitend veroorzaakt wordt

door rijping? We vermoeden dat de omvangrijke instructieteksten over de zes aspecten van teksten (Doelgerichtheid, Publiekgerichtheid, Opbouw, Betrouwbaarheid, Nieuwswaarde en Stijl) daartoe hebben bijgedragen, alsmede het lezen van elkaars teksten (herschreven versies) in beide condities. De nauwe aansluiting tussen de instructieteksten en de feedbackinstrumenten zorgde ervoor dat leerlingen de concepten Doel- en Publiekgerichtheid gingen begrijpen en trachtten heldere teksten te schrijven in het besef dat een tekst een transactioneel karakter heeft. We gaan op deze mogelijkheid hierna wat dieper in.

Het lijkt erop dat leerlingen vooral helderder zijn gaan schrijven en meer gebruik zijn gaan maken van taalmiddelen waarmee zij te kennen geven als schrijver te interacteren met een lezer. We presenteerden in de tabellen 7.10. en 7.11. de veranderingen tijdens het cursusjaar op analytische kenmerken van Doel- en Publiekgerichtheid. Doelgerichtheid valt uiteen in twee aspecten: Argumentatie en Samenhang (interne structuur). Op de meeste kenmerken van Doelgerichtheid was geen vooruitgang te constateren. Ook de scores verkregen met behulp van schaalbeoordeling wijzen op geen of een geringe vooruitgang. We concluderen dan ook dat Doelgerichtheid in de zin van Argumentatie niet, maar in de zin van Samenhang wel via ons onderwijsprogramma aan deze leeftijdsgroep onderwezen kan worden. Het kan zijn dat argumentatie niet via dit soort programma's onderwezen moet worden: het vereist een minder breed, specifieker programma wellicht. Het kan ook zijn dat leerlingen op deze leeftijd nog niet gevoelig zijn voor de kwaliteit van de argumentatie: veel opstellen bevatten wel een standpunt en een poging dat aannemelijk te maken, maar nog geen derde van die opstellen slaagt erin. We herinneren eraan dat leerlingen in het vooronderzoek te kennen gaven vooral de criteria over betrouwbaarheid moeilijk te vinden en die categorie als geheel het onbelangrijkst van de zes vonden. Dit betrouwbaarheidsaspect van Doelgerichtheid, dat betrekking heeft op argumentatieve teksten, is wellicht nog niet leerbaar. Een ander aspect van Doelgerichtheid, een heldere lijn in het informatie-aanbod, of 'de tekst moet ergens heen leiden' lijkt wel verworven te worden. (Vergelijk de resultaten op de variabele Opbouw en de procesvariabele Structureren). Opstellen worden wel meer 'reader based' (Flower, 1979). Leerlingen maken niet alleen teksten, die hun 'Knowledge Telling' model (Scardamalia en Bereiter, 1986) produceert, maar herordenen de informatie in een voor anderen heldere lijn. Ook de scores op een aantal kenmerken van Publiekgerichtheid (tabel 7.11) wijzen erop dat leerlingen de lezer gaan helpen de informatie te verwerken: het tekstthema wordt vaker geïdentificeerd in de inleiding en leerlingen gaan een inhoud- en/of structuuraanduidende titel gebruiken.

Uit de toename van de helderheid van opstellen, af te leiden uit scores op de analytische kenmerken van Doelgerichtheid en Publiekgerichtheid, veranderingen in schrijf-

processen en de oordelen op de variabele Opbouw, en misschien ook Stijl, en uit de toename van het gebruik van schrijver-lezer-interactie in inleidingen, concluderen we dat leerlingen elementen van de concepten Doel- en Publiekgerichtheid zich eigen zijn gaan maken. Een aantal van deze kenmerken komt rechtstreeks uit de instructieteksten: het gebruik van de tweede persoon enkelvoud en het stellen van een vraag bijvoorbeeld wordt sterk benadrukt in de instructietekst, evenals het gebruik van regels wit, inspringen, alinea's (Opbouw). Dergelijke kenmerken zijn voor een deel te benoemen als oppervlaktekenmerken. Als leerlingen die kenmerken één keer gedemonstreerd zien in de instructietekst of in een opstel van iemand anders, is het gemakkelijk die via imitatie in het eigen opstel op te nemen. Anderzijds gaat het om kenmerken die men niet tot oppervlaktekenmerken kan rekenen zoals coherentie (heldere lijn). In de instructietekst wordt er de nadruk op gelegd dat leerlingen moeten afleren hun eerste bedenksels als een opstel te zien. Ze moeten leren de inhoud voor lezers te herordenen. Ook de fase van herschrijven in de instructiecyclus maakt dat leerlingen het gewoon gaan vinden dat opstellen omgewerkt, veranderd, verbeterd kunnen worden. De vooruitgang in de schrijfprestaties schrijven we derhalve toe aan de instructietekst en de daarop toegesneden didactische maatregelen die leerlingen in staat stelden de inhoud ervan te verwerven en toe te passen. De keuze tussen twee van die didactische maatregelen, leerlingenrespons of docentfeedback, lijkt inhoudelijk niet doorslaggevend. Ze zijn equivalent en wellicht is de bijdrage van beide aan de leerwinst (zeer) gering.

## 8.3. Belangwekkende nevenresultaten

Opmerkelijk in ons onderzoek waren de lage correlaties tussen de scores op leerlingenniveau van verschillende tussentoetsmomenten op de Schrijfprestatievariabelen (tabel 7.1.). Zelfs al construeren we voor elke leerling een somscore voor zijn schrijfprestatie door zijn scores op de variabelen Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl te sommeren, dan nog is de stabiliteit van de leerlingscore laag: een correlatie van .38 tussen voortoets- en tussentoetsscores. Het betreft hier een echte stabiliteitsindex, want de voortoets- en de tussentoets waren identiek. De correlatie tussen voor- en natoets bedroeg .24; deze correlatie kan gedrukt worden doordat het onderwerp waarover geschreven moest worden verschilde. Vergeleken met voortoets-natoetscorrelaties uit in hoofdstuk 3 gerapporteerde effectstudies (Benson, 1973: .60; Ford, 1973: .92; Sager, 1973: .51) zijn onze correlaties laag. Een leerling die op de voortoets tot de voorhoede van de steekproef behoorde, leverde op de tussentoets een prestatie die hem achterin het peloton bracht. Zoiets kan gemakkelijk als de totale groep homogeen qua prestatieniveau is, zodat er

maar iets hoeft te gebeuren en het klassement ligt op zijn kop. De spreiding in de leerlingenscores is echter zo groot (de standaardafwijkingen liggen om en nabij de 10 punten) dat moeilijk van een voor schrijfprestaties homogene groep gesproken kan worden. Dat er een effect van de factor Onderwerp op de leerlingenprestatie is, wisten we uit eerder onderzoek (o.a. Wesdorp, 1974). Uit dergelijke onderzoeken blijkt dat een enkele opdracht geen goede indicatie van stelvaardigheid oplevert, omdat de prestaties nogal fluctureren. In ons onderzoek ging het er echter niet om de algemene schrijfvaardigheid te operationaliseren, maar een specifiekere, namelijk de prestatie voor betogende teksten. Door de opdrachten exact (voortoets en tussentoets) of nagenoeg hetzelfde (natoets) te formuleren, trachtten we het opdrachteffect te beperken. Als er sprake zou zijn van een onderwerpeffect zou dat binnen de condities voor dezelfde proportie leerlingen gunstig en ongunstig zijn, en zouden de verhoudingen op voor- en tussentoets dezelfde zijn. Uit de test-hertestcorrelatie van .38 blijkt dat we niet in onze opzet geslaagd zijn. Als het om individuele scores gaat, is het dus zelfs in die gevallen waarin men identieke opdrachten gebruikt bij het vaststellen van schrijfprestaties aan te bevelen verscheidene opdrachten van dezelfde soort voor te leggen om zo via testverlenging een betrouwbaardere index van de schrijfprestatie van een leerling te verkrijgen. Nu is het gebrek aan stabiliteit geen probleem in ons onderzoek, omdat wij analyseerden op klasseniveau. Door een klassescore te berekenen werd de leerlingfluctuatie als het ware uitgemiddeld. Dat de klassescore betrouwbaarder is, blijkt wel uit een vergelijking van de test-hertestcorrelaties van leerlingscores en klassescores (tabel 7.1.).

Voor de lespraktijk heeft deze bevinding verstrekkende gevolgen. Over het algemeen worden er in een cursusjaar niet veel opstellen geschreven. Onze in schrijfonderwijs geïnteresseerde docenten bijvoorbeeld zeiden gewoonlijk in de derde klas vier à vijf opstellen te laten schrijven; per rapportagekwartaal dus maximaal twee. Tussen de opdrachten bestaan dan nog verschillen in tekstsoort en onderwerp. Om op een dergelijk 'proefwerk' een rapportcijfer te baseren is erg riskant. Het geeft de leerling bovendien het idee dat schrijven iets is dat je in één keer moet kunnen bewijzen. Cijfers verkregen via opstellen zouden pas aan het eind van het cursusjaar mogen meetellen voor het rapport. Een tweede gevolg is dat men wat voorzichtiger zou moeten zijn in het onderwijs om zich uit te spreken over de onveranderlijkheid van de prestatierangorde in een klas. Alom heerst de opvatting dat schrijfonderwijs niet loont: de goede leerlingen blijven de goede, de slechte blijven relatief slecht. Docenten doen dergelijke uitspraken echter op basis van weinig waarnemingen en zullen waarschijnlijk ten prooi vallen aan de werking van het halo-effect en 'selffulfilling prophecy': de beoordeling van opstellen is over het algemeen zo onbetrouwbaar dat het algemene oordeel over de leerlingen een docent parten zal

spelen bij het vellen van een oordeel over diens stelvaardigheid, blijkend uit een enkel opstel.

Een tweede opmerkelijk verband troffen we aan tussen Psychologische Variabelen en Schrijfprestatievariabelen (tabel 7.11.). De correlaties waren veelal niet eens statistisch significant, hoewel een vrij grote steekproef (264) dat toch al snel mogelijk maakt. De hoogste correlatie was er een van .30 tussen Stijl en Houding-jegens-Schrijven. Een verklaring voor deze lage correlaties kan zijn dat de meetbetrouwbaarheid van beide variabelen te wensen overliet. Maar zelfs al construeren we stabiele scores, door per leerling de scores per Schrijfprestatievariabele en per Psychologische Variabele over de drie meetmomenten te sommeren, dan nog blijven de correlaties tussen Schrijfprestatievariabelen enerzijds en Psychologische Variabelen anderzijds laag. Ze variëren van -.05 (Doelgerichtheid en Houding-jegens-Beoordeeld-worden) tot .30 (Stijl en Angst-voor-het-niet-Kunnen-schrijven), met een gemiddelde van .15. Deze bevinding strookt met wat wij in enkele studies aantroffen. Benson (1979) en Pfeifer (1981) namen beide Daly en Millers (1975) Writing Apprehension Test af en lieten opstellen schrijven. De correlaties tussen schrijfangst en schrijfprestatie bedroegen respectievelijk .25 en .00. Ward (1959) vond een correlatie van .07 tussen opstelscores en scores op een instrument dat de attituden jegens de schrijfcursus operationaliseerde. Sears (1971) vond correlaties van .26 en .43 tussen opstelscores en respectievelijk de eigen schatting van inzet voor de schrijfcursus en de schatting van het eigen schrijfvermogen. Daly verrichtte geen correlatief onderzoek naar de relatie tussen schrijfangst en schrijfvaardigheid, maar stelde statistisch significante verschillen vast tussen sterk en weinig schrijfangstigen op diverse variabelen van een objectieve stelvaardigheidstoets (Daly, 1978). In een andere studie waarbij Daly betrokken was (Faigley et al., 1981), bleken hoog- en laagschrijfangstigen significant verschillend te presteren op verhalende stelopdrachten. Het verschil trad echter niet op bij argumentatieve stelopdrachten.

In kringen van vakdidactici Nederlands en curriculumdeskundigen voor het vak Nederlands heerst de opvatting dat in het schrijfonderwijs het houdingsaspect beklemtoond dient te worden. In een advies aan de minister over de inhoud van het Voortgezet Basisonderwijs wordt zelfs door S.L.O.-medewerkers geschreven dat het weinig zin heeft als leerlingen iets kunnen (n.l. schrijven), maar het niet willen. De teneur is: als leerlingen schrijven maar leuk vinden, dan komt dat kunnen min of meer vanzelf wel. Als er echter zo'n zwak of geen verband is tussen houdingen en schrijfprestaties zou men toch het accent in het onderwijs eerder op het bevorderen van de prestaties dan op de attituden moeten leggen. Een zwakke samenhang betekent overigens niet, dat er geen leerlingen zijn, bij wie het verband tussen Psychologische Variabelen en Schrijfprestaties niet sterk is, en die wellicht gebaat zouden

zijn bij een attitudeveranderend onderwijs. Echter, de implicatie hiervan is dat men deze leerlingen moet leren onderscheiden van anderen om remediërend onderwijs te kunnen geven. Cognitief gericht schrijfvaardigheidsonderwijs moet men echter niet bij voorbaat afwijzen, omdat het het schrijfplezier niet zou bevorderen en daardoor ineffectief onderwijs zou zijn: voor velen doet het er niet toe of zij schrijven nou leuk vinden of niet om goed te kunnen presteren.

We moeten erop verdacht zijn, dat er variabelen in het spel kunnen zijn die de relatie tussen Psychologische Variabelen en Schrijfprestatievariabelen maskeren. Pfeifer (1981) vond correlaties tussen extraversie, introversie, schrijfangst en schrijfprestaties, die aanwijzingen bevatten dat een zekere mate van schrijfangst zoals gemeten met de Writing Apprehension Test (Miller en Daly, 1975) voor extraverte personen faciliterend kan werken op de schrijfprestatie, terwijl een zelfde niveau van schrijfangst voor introverte leerlingen remmend werkte op de schrijfprestatie. Verder onderzoek naar verbanden tussen (componenten van) schrijfangst en schrijfprestaties zou meer persoonlijkheidskenmerken in het ontwerp moeten opnemen. Een van de relevante persoonlijkheidskenmerken zou attributievoorkeur kunnen zijn. De lage correlaties tussen de variabele Angst-voor-het-niet-Kunnen-schrijven en schrijfprestaties betekenen immers dat er heel wat leerlingen zijn die zich niet terecht zorgen maken over hun prestaties: hoewel zij relatief goed presteren, maken zij zich toch zorgen over het resultaat van hun stelpogingen. Een gezonde dosis relativeringsvermogen kan geen kwaad, maar anderzijds lijkt het verkieslijker als leerlingen met een relatief goede prestatie die prestatie aan eigen kunnen toeschrijven. Interne attribuering is volgens attributietheoretici bij goede prestaties bevorderlijk voor het leveren van volgende goede prestaties. Naast persoonlijkheidskenmerken als intro- en extraversie zou schrijfangst- en attitudenonderzoek daarom attributievoorkeuren kunnen opnemen als variabele.

## 8.4. Consequenties voor de lespraktijk

Van de twee onderwijsprogramma's die wij beproefden, stelden we vast dat zij geen verschillen in effecten teweeg brachten op schrijfprestaties en attituden en dat beide positieve effecten bewerkstelligden op schrijfprocessen. Verder reikte onze onderzoeksvraag niet. Aan het slot van deze studie willen we echter terugkeren bij waar we in de inleiding begonnen: de praktijk van het stelonderwijs, de docent. Voor welke van beide programma's kiest hij? Is er sprake van een doelpuntloos gelijk spel of hebben beide programma's gescoord?

De eerste stap in zo'n keuzeproces is af te wegen of het spel correct is gespeeld. Is er risico genomen? Is er van beide kanten aanvallend gespeeld of bleef één partij voortdurend met elf man in het eigen doelgebied? Wij vinden dat in

dit onderzoek inderdaad twee methoden op het scherp van de snede, aanvallend hebben gestreden. Het ging hier niet, - wat in effectonderzoeken wel gebeurt -, om een vergelijking tussen een gedegen ontwikkeld experimenteel programma en een controleprogramma dat weinig voorstelt. Hier ging het om twee aan elkaar gewaagde elftallen.

De tweede stap is: hebben beide programma's gescoord? Hoewel in dit onderzoek geen controle voor groei is opgenomen en er strikt genomen geen antwoord op deze vraag gegeven kan worden, wijzen we toch op enkele gegevens die bijdragen aan een antwoord. Op de Schrijfprestatievariabelen kan een vooruitgang vastgesteld worden, als men aanneemt dat de verschillen tussen de voor- en tussentoets niet alleen veroorzaakt worden doordat dezelfde stelopdracht voorlag en de verschillen tussen de tussen- en de natoets niet alleen veroorzaakt worden doordat de opdrachten verschilden. De bijdrage van het onderwijsprogramma op die vooruitgang kan op drie manieren beargumenteerd worden. Ten eerste: voor enkele variabelen, Stijl en Opbouw, is die groei zo spectaculair te noemen, dat er bij een dergelijke gestage groei in volgende cursusjaren weinig meer te klagen zal zijn over de schrijfvaardigheid van schoolverlaters. Ten tweede: uit de gegevens over de kenmerken van inleidingen wat de Publiekgerichtheid betreft, blijkt dat leerlingen kenmerken in hun inleidingen verwerken die typisch uit het onderwijsprogramma voortkomen, zoals het gebruik van je en jij in de tekst om een interactie met de lezer tot stand te brengen. Ten derde: collega's van docenten die aan het onderzoek meewerkten, die in de vierde klassen leerlingengroepen onderwezen, die voor een gedeelte uit aan het onderzoek deelnemende leerlingen bestonden, waren zeer te spreken over de prestaties van deze leerlingen. Zij onderscheidden zich duidelijk van vierdeklassers die niet in het derde leerjaar aan het onderzoek meewerkten.

Als beide programma's leiden tot aanvaardbare resultaten, lossen zij dan ook praktijkproblemen op? In de inleiding van deze studie haalden wij een onderzoek van Damhuis, De Glopper en Wesdorp (1983) aan, waarin onder andere nagegaan is wat volgens docenten Nederlands problemen in het schrijfonderwijs zijn. We memoreerden er drie: tijdgebrek, beoordelingsproblemen en kwaliteit van de opstellen, vooral wat de inhoud en de stijl betreft. Leveren onze programma's een bijdrage aan het oplossen van een of meer van deze drie problemen? Beide programma's vergen veel lesuren. In de 22 klassen in ons onderzoek domineerde het schrijfonderwijs als vakonderdeel zozeer, dat sommige van de betrokken docenten vreesden hun gewone programma voor het derde leerjaar niet af te kunnen krijgen. Wat ons betreft wordt schrijfvaardigheid de spil van het curriculum Nederlands en worden andere vakonderdelen ten dienste van het schrijfonderwijs gesteld. Lezen, luisteren en spreken kunnen steeds functioneel beoefend worden in het kader van informatieverwerving over het onderwerp waarover geschreven moet worden. In deze zin kunnen onze programma's gezien

worden als een nadere uitwerking van het S.L.O.-project Gericht Schrijven (De Zanger, 1981). Wij zouden willen dat de schrijfvaardigheid een hogere prioriteit in het H.A.V.O. en het V.W.O. krijgt. Een docent die klaagt over tijdgebrek in het schrijfonderwijs, bedoelt daarmee blijkbaar aan andere vakonderdelen evenveel of meer waarde te hechten. Want zoveel tijd besteedt men gemiddeld niet aan stelonderwijs (Damhuis et al., 1983). Zo'n klacht maakt duidelijk waar in het vak Nederlands de schoen wringt. Alles kan bij het vak, alles is taal, alles is communicatie. Een treurig voorbeeld van een verworden vakopvatting is het al eerder genoemde advies van de S.L.O. aan de minister over de inhoud van het vak Nederlands in het Voortgezet Basisonderwijs. Het zou goed zijn als het schoolvak Nederlands een duidelijke identiteit krijgt door een transparant, op vaardigheden en prestaties gericht curriculum. Verzoeken om meer lesuren Nederlands zullen eerder gehoor vinden als men op basis van programma's en resultaten kan laten zien dat er onderwijs gegeven wordt, dan wanneer men een zoveelste vakdidactische nouveauté aan de verzameling 'inhoudselementen' toevoegt. Tijdgebrek, kortom, is veel minder een probleem als er keuzes worden gemaakt.

In onze programma's is het beoordelen, afgezien van de tijd die het kost, niet zo'n groot probleem meer. Instructie en beoordeling zijn nauw op elkaar afgestemd. Beoordelingscriteria geven de beoordelende docent of medeleerling steun. We onderkennen wel dat veel docenten in hun lespraktijk geen half uur aan een opstel kunnen besteden zoals de docenten dat in ons onderzoek deden. Leerlingenrespons is dan een alternatief. Maar waarschijnlijk kan leerlingenrespons niet door iedere docent en niet in iedere klas geïmplementeerd worden. Werken met leerlingenrespons in onze opzet vergt nogal wat organisatie en planning van de docent. Alle opstellen moeten op een bepaald moment beschikbaar zijn om te kopiëren voor commentatoren. Alle commentaren moeten op een zeker ogenblik beschikbaar zijn voor herschrijvers. Er moeten maatregelen getroffen worden waardoor leerlingen absente medeleerlingen toch in de lessencyclus betrekken. We zagen dat het argumenteren in teksten derde klassers nog niet goed afgaat. Zo zullen er in andere leerjaren andere subvaardigheden zijn die te hoog gegrepen zijn. Verder zou het kunnen zijn dat een klas aan zekere voorwaarden op sociaal-psychologisch gebied als motivatie, coöperatie en cohesie moet voldoen om met vrucht de werkvorm leerlingenrespons te gebruiken. Wij denken overigens dat dat wel mee zal vallen: in die klassen waarvan de bij het onderzoek betrokken docenten zeiden aan het begin van het cursusjaar dat ze daarin liever docentfeedback hadden gegeven dan leerlingenrespons, bleek leerlingenrespons toch ook te lukken.

Het laatste praktijkprobleem is de kwaliteit van de opstellen. Schrijfonderwijs in het voortgezet onderwijs is net als schrijven zelf vaak een vrij onbeheersbaar onderwijs. Zicht op vooruitgang, op effect van het onderwijs, is er niet, hetgeen

wellicht leidt tot eerder genoemde opmerking dat schrijfonderwijs niets uitricht (par. 8.3.). De verschillen tussen de onderzoeksopstellen, geschreven in augustus van het onderzoeksjaar en die van mei, zijn duidelijk. Er is vooruitgang geconstateerd. De docenten die bij het onderzoek betrokken waren die de werkmappen van leerlingen na elke schrijfronde nakeken, bemerkten bij leerlingen vooruitgang, juist ook omdat de opstellen van verschillende schrijfronden zich lieten vergelijken omdat zij onderdeel waren van een en dezelfde instructie en behoorden tot één en dezelfde tekstsoort. Die vergelijkbaarheid achten wij een groot goed om zo meer greep op het onderwijsleerproces en de vorderingen van leerlingen te krijgen. Leerlingen zelf bemerkten hun vooruitgang ook. Als zij in een trimester de keuze hadden uit verschillende opstellen om de docent voor te leggen voor een cijfer, meetellend voor het rapport, kozen zij veelal het laatst geschreven opstel. Hoewel er zeker nog aan de beide onderwijsprogramma's verbeteringen aangebracht kunnen worden, menen wij dat zij zullen leiden tot kwalitatief aanvaardbare opstellen.

Gezien de in de inleiding gememoreerde praktijkproblemen van tijd, beoordelen en kwaliteit die we in deze evaluatie bespraken, valt de keus tussen beide programma's zoals in het onderzoek uitgevoerd, niet zwaar. Doorslaggevende factor is dan de tijd: weinig docenten zullen vier maal per jaar een halfuur aan een opstel van een leerling kunnen besteden. We achten het echter niet uitgesloten (par. 8.2.) dat veel minder feedback tot dezelfde resultaten zou leiden. In dat geval zou de keuze tussen leerlingenrespons en docentfeedback afhankelijk zijn van de vaardigheden van de docent en eigenschappen van de klas, alsmede van de rol in het onderwijsleerproces die men aan leerlingen toekent.

SUMMARY

Does a programme of written composition teaching based on peer feedback lead to better results than one based on teacher feedback? In modern mother-tongue curricula, educational publications and handbooks and publications in which teachers describe and promote their own methods, peer evaluation, peer feedback or peer teaching is accorded an important role in the teaching of written composition skills. Two schools of thought in thinking about mother-tongue curricula have led to heightened attention to peer feedback as an educational technique. The first is a more general educational movement which had redefined the role of the pupil in the teaching/learning process. Teaching is there no longer seen as a purely transactional process between teacher and pupil, but rather as an interpretative process in which the pupil's knowledge is itself evaluated and used as a component of the teaching/learning process. The second movement is specific to the teaching of language: the communicative shift. Here it is stressed that pupils learn from valid feedback, i.e. the reactions of the intended readers. Written composition tasks are formulated in such a manner that the audience is composed of pupils. In such a class, pupils constitute communicative pairs (writers and readers) and instructive pairs (feedback givers and receivers).

On the basis of an analysis of literature on peer evaluation and written composition, written composition processes and educational theory, three groups of dependent variables were defined by which pupils ought to differ in consequence of the experimental variable.

a) It was expected that experimental classes would perform better than control classes on the writing achievement variables of Goal Orientation, Audience Orientation, Organization and Style.
b) It was expected that experimental classes would display a lower level of Fear-of-not-being-able-to-write -- the cognitive component of writing comprehension -- and a more positive Attitude-towards-being-evaluated and Attitude-towards-writing than control classes.
c) It was expected that writing processes in the experimental classes would differ from those in the control classes. Pupils in the experimental classes were expected to perform more planning activities such as self-instruction, generation of ideas, criteria and presentation structures, and structuring of ideas. The formulation process was expected to be more frequently interrupted by a choice process in which available words and phrases were weighed up against each other. The revision process would then be characterized by more frequent evaluation and revision of textual elements and longer passages. Finally pupils in the experimental condition were expected to be characterized by a higher frequency of interactional monitoring: there would be both 'feedforward' in the generation and structuring and 'feedback' in the evaluation and revision of text.

To determine the effect of peer evaluation on these different aspects of written composition skill, an experiment was carried out consisting of three parts: development of an experimental teaching programme, construction of the measuring instruments, and execution and monitoring of a field experiment.

Development of an experimental teaching programme.
At a school the author and two fellow teachers developed and tested a procedure for teaching the writing of goal-oriented and audience-oriented discursive texts in four third forms in two sorts of secondary school. By means of process research, questionnaires, analysis of the reliability of the feedback instruments, and learner reports, this procedure was evaluated and refined. The final programme contained four block courses each of 10-12 lessons. All block courses were constructed on the same pattern.

A. Preparatory lessons (3-6). In these lessons the pupils studied instruction texts on aspects of writing and texts. Two aspects were introduced in each part course: goal and audience orientation in the first course, organization and reliability in the second, news value and usage in the third. In the fourth course the instruction text contained a synthesis of all the steps in the process that had been dealt with up to that point. Also in the preparatory lessons, some tasks were used to study writing tasks and information on the subject consisting of articles or cuttings from newspapers, magazines and brochures. The pupils then wrote, in class, first a rough draft and then a first version. As a reflection task they then described how their first version had been written ('How did the writing process go this time?') and what they themselves thought of their texts ('What do I think of my essay?').

B. Commentary lessons. Each pupil was given by the teacher a copy of the essays of three arbitrarily selected, anonymous peers. He read each essay and by means of a subjective reaction form with items like 'I am not convinced' indicated his impression of what he had read. After this first reading the pupil carried out his second reflection task: 'What have I learned by reading other people's texts?' Then he read the same essays more closely and gave more detailed feedback by answering questions from a comment questionnaire. The aspects on which comments were given corresponded to those in the instruction texts. Five comment questions were asked on each aspect. Besides answering these comment questions the commentator also had to carry out a variety of tasks in the essays themselves, such as indicating audience-oriented sentences and phrases and the transitions between beginning, middle and end. At the same time some of the comment questions obliged the commentator to indicate features to be judged positively or negatively in the essay itself.

Commenting on an essay took 30-45 minutes. Pupils performed the task in two lessons; what they did not finish in class

they did at home. For the final conclusion of this stage pupils carried out a third reflection task: 'What have I learned from commenting on other people's texts?'

C. Processing comments. Pupils were given subjective reader-based and objective criterion-based feedback, all notes on their essays, and any further written notes to these assessments, from three anonymous commentators. They then ordered the comments, summarized them and wrote comments on them. It was stressed in their instructions that the writers were responsible for their own choice from the feedback, but that they must give reasons for their choice. The pupils then draw up a rewriting plan in which they indicated what they planned to chance and how they wanted to change it. Processing comments was done partly at school (2-3 lessons) and partly at home.

D. Final version. On the basis of the rewriting plan a final version was written in the course of a single lesson.

The pupil then gathered his work together in a ring binder. The teacher evaluated the binders for each block course for their completeness, neatness and the quality of the work, with the exception of the essays. Every quarter a pupil selected one of his essays for evaluation by the teacher for a mark counting towards his end-of-term report.

The pupils in the control groups went through exactly the same programme, except that at stage B it was the teacher who provided the feedback, using the same forms, criteria and tasks as in the experimental condition. The teacher spent approximately half an hour on each essay.

Instrumentation.

Essay scales were constructed to measure writing performance variables (Goal Orientation, Audience Orientation, Organization and Style). These were a series of essays of gradually increasing quality which are a useful aid to raters because differences in quality and the associated textual features are clearly shown. In the case of Goal Orientation and Audience Orientation analytical scoring rules were also drawn up.

To measure writing process variables a detailed coding scheme was developed on the basis of the writing process model of Hayes and Flower, the hypotheses, and a number of test protocols.

For the Psychological Variables an inventory was constructed on the basis of Miller and Daly's Writing Apprehension Test and Bergen's Situation-Specific Apprehension Test.

The reliabilities of the essay scales in terms of rater reliability were good: the correlation between two raters varied from .64 to .80; intra-reliability was even higher. The test reliability of the three Psychological Scales varied from .80 to .93 (Cronbach's alpha). Some items out of the analytical scoring rules for Goal Orientation and Audience Orientation proved still to be insufficiently objectively scorable.

Experiment.
Eleven experienced teachers at eight different schools of two different levels of secondary school carried out the programme in the third form. Each teacher taught two classes randomly assigned to the experimental or control condition. Before, during and after the course the pupils wrote a discursive essay, for research purposes, that had no further function in the teaching programme. The tasks for the pre-test and intermediate test were identical; that for the post-test was similar. At the three test sessions the pupils' attitude inventories were also taken. During the pre-test and intermediate test periods writing aloud protocols were elicited from two pupils. The implementation of the teaching programmes was monitored with teacher and pupil log-books, observations and questionnaires.

In each class a random sample was taken of all pupils for whom three test essays were available. The 792 essays (22 classes x 12 pupils x 3 tests) were typed out and evaluated by two raters using the essay scales for Goal Orientation, Audience Orientation, Organization and Style developed during the project. The essays were also scored according to scoring rules for Goal Orientation and Audience Orientation. One of the two writing aloud protocols from each class was typed out according to the set procedure. These 44 protocols were coded by means of a fine-meshed scheme.

The class scores on Writing Performance Variables and Psychological Variables were analysed by multivariate co-variance analysis. The Writing Process Variables were analysed by loglineair model-fitting.

The results were fairly unambiguous. The difference in the teaching programme produced no differences on Writing Performance and Psychological Variables. On the analytically scored items for Goal Orientation and Audience Orientation there proved to be no difference between the conditions, though there was an increase in the number of essays that followed a clear line of thinking and displayed some form of interaction with the reader. In both conditions the scores for all variables were clearly higher on the intermediate test and post-test than on the pre-test, with the exception of the scores for Audience Orientation variable in the post-test. Indeed, perhaps because of a task effect, these scores were even lower than that for the pre-test. There were also some differences in the Writing Process Variables. In the experimental condition the number of self-instructions rose, in the control condition it fell; in the control condition the number of times increased that pupils posed themselves a choice in the matter of wording; in the experimental condition it fell. In both conditions there was less generation in the post-test but more structuring and more frequent alteration of words and phrases.

Some explorations were carried out which showed that the Teacher factor had an effect on almost all Writing Performance

Variables in the intermediate test and post-test. It also emerged that the Sex factor had some effect, but no consistent pattern could be demonstrated of any statistically significant main effects of sex or of sex/programme interaction effects. From some correlation calculations it emerged that the Psychological Variables bear little or no relation to written composition performance during the three tests: a maximum of 9% of the variance of the scores in the Style category is explained by scores on Psychological Variables. Pupil Writing Performance scores proved unstable: correlations between pre-test and intermediate test scores were low.

In the discussion an attempt is made to explain the results of this research. Perhaps pupils do learn from giving feedback, but perhaps the feedback they receive from their peers is of a lower quality that that received from teachers. However, a check carried out with questionnaires showed that in both conditions few pupils were dissatisfied with the clarity, carefulness and helpfulness of the feedback they received. Perhaps - another explanation - feedback is not really all that important: there proved to be no significant correlation between written composition scores and dissatisfaction with feedback. A third possible explanation for the absence of differences is that in the experimental condition pupils acquire criteria for good texts but that this knowledge does not lead directly to their application in writing tasks.

Assuming that the observed improvement in written composition skill cannot be attributed solely to cognitive growth, the conclusion is that peer feedback and teacher feedback are in equal competition. The choice between the two written composition teaching programmes tried here must depend on the skills of the teacher concerned, the time he can and will devote to essay correction, and the degree to which the pupils may be considered able to meet the conditions for the proper functioning of peer evaluation. Intensive peer evaluation is a good, but not a better, alternative to intensive teacher feedback.

LITERATUUR

Amabile, T.M. en A.H.Glazenbrook, A negativity bias in interpersonal evaluation. In: Journal of experimental social psychology, jrg.18, no.1, 1982, 1-22.

Baltzer, J., Hardop denken bij het schrijven van een opstel. Een onderzoek naar effecten van twee steldidactieken op schrijfprocessen van 15-jarigen, m.b.v. een analyse-schema voor hardop-denkprotocollen. SCO-cahier, S.C.O. Amsterdam, 1986.

Baltzer, J. en I. Bochardt, Het analyseren van hardop-schrijfprotocollen. Achtergronden en instructie. S.C.O. Amsterdam 1985.

Bandura, A., Self efficacy mechanisms in human agency. In: American Psychologist, jrg.37, no.2, p.122-147.

Barry, L., Dr. Strangelove attends the Iowa Institute on Writing, or to kill the 500 word theme and learn to love Composition. Paper presented at the Annual Meeting of the Northwest Regional Conference of the National Council of Teachers of English (April 10-12). Portland, 1980. (ED 186 920).

Bata, E.J., A study of the Relative Effectiveness of Marking Technique on Junior College Freshman English Composition. Diss. Univ. of Maryland. Z.p., 1972.

Bean, J.C., A method of peer-evaluation of student writing. In: College Composition and Communication, jrg.30, no.3, 1979, p. 301-302.

Beaven, M.H., Individualized goal setting, self-evaluation peer evaluation. In: C.R.Cooper en L.Odell, Evaluating Writing: describing, measuring, judging. Z.p. (Urbana), 1977, p.135-156.

Beck, P., Peer Tutoring at a Community College. In: College English, jrg.40, no.4, 1978, p.437-439.

Bell, E.S., The peer tutor as principal benefactor in the writing center or it's not just for english teaching anymore. Paper Meeting of the Southeastern Writing Center Conference, Columbia, 1983. (ED 234 407).

Benesch, S., Improving peer response: collaboration between teachers and students. Paper Annual meeting of the Conference on College Composition and Communication. New York, 1984. (ED 243 113).

Bennett, E.M., R.L. Blomquist and A.C. Goldstein, Communication through limited response questioning. In: Public Opinion Quarterly, jrg.18, 1954, p.303-308.

Benson, N.L., The effects of peer feedback during the writing process on writing performance, revision behavior, and attitude toward writing. Diss. University of Colorado. Boulder, 1979.

Bergen, Th.C.M., Evaluatie-angst en vermijdingstendens. Een onderzoek naar de oriëntatie van leerlingen om mislukking te vermijden in taaksituaties tijdens de les. Diss. K.U.N. (S.V.O.-reeks no.47). S.V.O. 's-Gravenhage, 1981.

Bergh, H. van den, et al., Instrumenten ter peiling van de taalvaardigheid van leerlingen uit de derde klas v.o. S.C.O.-rapport 71. S.C.O. Amsterdam, 1986.

Bergh, H. van den, Bijlagen bij rapport Instrumenten ter peiling van de taalvaardigheid van leerlingen uit de derde klas v.o.. S.C.O.-rapport 71. S.C.O. Amsterdam, 1986.

Berkenkotter, C., Student Writers and their Audience: Case Studies of the Revising Decisions of Three College Freshmen. Paper presented at the Annual Meeting of the Conference on College Composition and Communication (March). Detroit, 1983. (ED 236 618).

Berkenkotter, C., Student writers and their sense of authority over texts. In: College Composition and Communication, jrg.35, no.3, (Oct.), 1984, p. 312-319.

Beyer, B.K. and A.Brostoff, The time it takes: Managing/Evaluating writing and social studies. In: Social Education, jrg.43, no.3, 1979, p.194-197.

Biegstraaten, J. e.a., Functioneel Nederlands, (3VH). Synthetische methode Nederlands voor vwo, havo en mavo. Wolters-Noordhoff. Groningen, 1977.

Bissland, J.H., Peer Evaluation method promotes sharper writing. In: Journalism Educator, jrg.34, no.4, 1980, p.17-19.

Blake, R.W. and F.B.Tuttle, Composing as the curriculum: The Albion Writing Project. Program prepared at State University of New York College at Brockport. Z.p., 1977. (ED 146590).

Blok, H. & J.B. Hoeksma, Opstellen geschaald. De constructie van beoordelingsschalen voor vijf schrijfopdrachten. S.C.O.-rapport 45. S.C.O. Amsterdam, 1984.

Blok, H. en G. Rijlaarsdam, Beoordeling van schrijfprodukten door leerlingen: evaluatie van de kwaliteit. In: Levende talen, nr.367 (dec.), 1981, p.958-972.

Bloom, B.G., Human Characteristics and School Learning. McGraw Hill. New York, 1976.

Bloom, S., Peer and cross-age tutoring in the schools. An individualized supplement to group instruction. U.S. Department of Health, Education, and Welfare. Washington, 1976.

Bochardt, I. , Het schrijfproces: cognitief-psychologisch onderzoek van Flower and Hayes. In: Tijdschrift voor taalbeheersing, jrg.6, no.1, (feb.), 1984, p.23-42.

Bochardt, I., Onderzoek naar het schrijf- en planningsproces in de zeventiger en begin tachtiger jaren, op weg naar een theorie. Een literatuurverslag en een verslag van een hard-op-denk-experiment. S.C.O. Amsterdam, 1983.

Bochardt, I. & G. Rijlaarsdam, Reageren op elkaars opstellen. Een onderzoek naar de bruikbaarheid van een didactisch beoordelingsinstrument. Projectpublikatie 2 van Z.W.O.-projecten 'Onderzoek naar effecten van beoordeling door medeleerlingen op aspecten van de stelvaardigheid'. S.C.O. Amsterdam, 1984.

Bonset, H., De relatie onderzoeker-leerkracht in etnografisch en case-study-onderzoek. In: L. Lentz, J. Sturm en H. van Tuijl (red.), Schrijfonderwijs onder het mes. S.L.O. Enschede, 1986.

Booher, S.C., A report on the tutorial outreach model for reading and writing across the curriculum at Los Medanos College. Pittsburg, 1982 (ED 221 252).

Book, C. and K. Wynkoop Simmons, Dimensions and perceived helpfullness of student speech criticism. In: Communication Education, jrg.29, no.2, (May), 1980, p.135-145.

Bouton, K. and G.Tutty, The effect of peer-evaluated student compositons on writing improvement. In: The English Record, 26, 1975, p.64-69.

Bradley, J.V., Pernicious publication practices. In: Bulletin of the Psychonomic Society, jrg.18, no.1, 1981, 31-34.

Braet, A., De inleiding en het slot van affirmatieve teksten. Een voorbeeld van klassiek-retorische communicatie-analyse. In: Levende Talen, nr.331, (mei) 1979, p.161-174.

Brennan, R.L. & D.J. Prediger, Coefficient kappa: some uses, misuses, and alternatives. In: Educational and psychological measurement, jrg.41, 1981, p.687-699.

Brinke, S. ten, The complete mother-tongue curriculum. A tentative survey of all the relevant ways of teaching the mother tongue in secondary education. Wolters-Noordhoff-Longman. Z.p., z.j. (Groningen, 1976).

Bruffee, K.A., A short course in writing. Practical Rhetoric for Composition Courses, Writing Workshops, and Tutor Training Programs. Little, Brown and Company. Boston, z.j. (1980).

Bruffee, K.A., The structure of kwowledge and the future of liberal education. In: Liberal Education, 1981, (Fall), 181-185.

Bult, E.H., E.S.F. Feteris, N.Lewis, Opbouw. (Deel 3 HAVO/VWO) Stijl, grammatica en spelling voor de derde klassen havo en vwo. Meulenhoff Educatief. Amsterdam, z.j. (1983).

Burt, R.M., Effects of an individualized, humanistic program of confluent literature and composition instruction on the writing performance and low ability suburban eleventh grade High School students. Diss. Temple University. Z.p., 1980. (U.M.I. 80-25 181).

Buys, A., Reflecteren op het opstel. In: Moer, nr.3, 1984, p.2-12.

Calkins, L., Writers need readers, not robins. In: Language Arts, jrg.55, no.6, 1978, p.704-707.

Calzonetti, F.J., Group review approach to the teaching of energy geography. Paper presented at the Annual Meeting of the National Council for Geographic Education (October 29). Pittsburg, 1981. (ED 212 526).

Camplese. D.A. and J.A.Mayo, How to improve the quality of student writing: the colleague swap. In: Teaching of Psychology, jrg.9, no.2, 1982, p.122-123.

Carter, R.D., By itself one-shot peer group revision has no power. Revised version paper delivered at National Council of Teachers of English. (Nov.) Washington, 1982. (ED 226 350).

Cate, Th.J. ten, Th.J.M. Tromp en M.G. Cornawall, De student als docent. (Aula 814 . Het Spectrum. Utrecht, 1984.

Cederblom, D. and J.W. Lounsbury, An investigation of user acceptance of peer evaluations. In: Personnel Psychology, jrg.33, no.3, 1980, p.567-579.

Christensen, J., Let's write it right. A student-oriented approach for teaching letterwriting skills. Paper presented at the 42nd Annual and 3rd International Convention of the American Business Communication Association. (December 28-30). Chicago, 1977. (ED 161 068).

Clifford, J., Composing in stages: The effects of a collaborative pedagogy. In: Research in the teaching of English, jrg. 15, 1981, p. 37-53.

Clifton, L.J., What if the kids did it? Paper presented at the Annual Meeting of the Northwest Regional Conference of the National Council of Teachers of English (Portland, OR, April 10-12). Portland, 1980. (ED 186 945).

Cohen, J., Statistical power analysis for the behavioral sciences. (Rev.ed.). Academic Press. New York, 1977.

Coleman, V.J., Competency-Based Tests: What are the alternatives? Paper presented at the Annual Meeting of the Conference on College Composition and Communication (29th, March 30-April 1). Denver, 1978. (ED 162 313).

Collins, J.L., Establishing peer evaluation of writing: students need an informed teacher model. Z.p., 1983. (ED 243 122).

Commissie Modernisering Moedertaalonderwijs, Advies over het eindexamen Nederlands in het V.W.O., het H.A.V.O. en het M.A.V.O. Staatsuitgeverij. Z.p. ('s-Gravenhage), 1977.

Cook, T.D. en D.T. Campbell, Quasi-Experimentation. Design and analysis Issues for field settings. Houghton Mifflin Company. Boston, z.j. (1979).

Copland, J.B., A twelve week study on the effects of peer evaluation for improving narrative writing performance at the eight grade level. Diss. Auburn University. Auburn, 1980.

Covington, D.H., A technical communication course using peer evaluation of writing. In: Engineering Education, jrg.69, no.5, 1979, p.417-419.

Crews, R., Teaching writing as an extension of the directed reading activity. Paper Annual Meeting of the Florida Reading Association. Hollywood, 1983. (ED 239 242).

Crowhurst, M., The writing workshop: An experiment in peer respons to writing. In: Language Arts, jrg.56, nr.7, 1979, p.757-762.

Daly, J.A., Writing apprehension and writing competency. In: Journal of educational research, jrg. 72, (sept.-oct.), 1978, p. 10-14.

Damhuis, R., K.de Glopper en H.Wesdorp, Het opstelonderwijs. Stelvaardigheid in het voortgezet onderwijs: theorie en praktijk. S.C.O. rapport, S.C.O. Amsterdam, 1983.

Danis, F., Weaving the web of meaning: Interaction patterns in peer-respons groups. Paper presented at the Annual Meeting of the Conference on College Composition and Communication (33rd, March 18-20). San Francisco, 1982. (ED 214202).

Delaney, M.C., A comparison of a student-centered, free writing program with a teacher-centered rhetorical approach to teaching college composition. Diss. Temple University. Z.p., 1980.

Dis, L.M. van (red.), Didactische handleiding voor de leraar in de moedertaal. Meulenhoff, Muusses, Wolters-Noordhoff. Amsterdam, Purmerend, Groningen, z.j.

Dollard, J. and N. Miller, Personality and psychotherapy. McGraw-Hill. New York, 1950.

Drop, W., De centrale betekenis van doelstelling en publiekgerichtheid bij het schrijven van zakelijke opstellen. In: Levende Talen, nr.275 (feb.) 1971, p. 77-83.

Dweck, C.S., W. Davidson, S. Nelson and E. Bradley, Sex differences in learned helplessnees: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. In: Developmental Psychology, jrg.14, no.3, 1978, p.268-276.

Earls, T.D., Peer evaluation and teacher evaluation of the first drafts of high school sophomores compositions. Diss. Boston University. Boston, 1983. (UMI 83-19886).

Eigen zorg, Het alternatieve opstel. Verslag van een experiment met de stelopdracht b-2 aan vier scholen in het jaar 1970/1971. Samengesteld door de vakgroep begeleiding experimenten Nederlands 'Eigen zorg'. Commissie vwo-havo-mavo. Staatsuitgeverij. Den Haag, 1971.

Eigen zorg, Anders opstellen. Verslag van het experiment met de stelopdracht b-2 aan tien vwo-scholen in het jaar 1971/1972. Samengesteld door de vakgroep begeleiding vwo-experiment Nederlands 'Eigen zorg'. Commissie vwo-havo-mavo. Staatsuitgeverij. Den Haag, 1972.

Eigen zorg, Meer dan opstellen. Eindverslag van het experiment met de stelopdracht b-2 aan vijftien vwo-scholen in de jaren 1970/1973. Samengesteld door de vakgroep begeleiding vwo-experiment Nederlands 'Eigen zorg'. Commissie vwo-havo-mavo. Staatsuitgeverij. Den Haag, 1974.

Elbow, P., Writing without teachers. Oxford University Press. New York, 1974.

Elbow, P., Writing with power. Techniques for mastering the writing process. Oxford University Press. New York, 1981.

Elias, K.M., Peer interaction: a method of creating voice in writing. University of Connecticut. Z.p., 1982. (ED 216 356).

Ellman, N., Peer evaluation and peer grading. In: English Journal, jrg.64, no.3, 1975, p.79-80.

Engel, M.F. and T.M. Sawyer, Contractual Revision. Paper Annual Meeting of the Conference of College Composition and Communication (March). Detroit, 1983.

Erftemeijer, A.G.I., Stijloefening op de middelbare school. Diss. K.U.N. Nijmegen, z.j. (1958).

Ericsson, K.A. and H.A. Simson, Protocol analysis. Verbal reports as data. MIT Press. Cambridge (Ma), 1984.

Er wordt steeds meer fout geschrefen. Verkennende nota over het moedertaalonderwijs. Staatsuitgeverij. 's-Gravenhage, 1981.

Faigley, L., J.A. Daly and S.P. Witte, The role of writing apprehension in writing performance and competence. In: Journal of educational research, jrg.75, no.1, (sept.-oct.), 1981, p.16-21.

Farrell, K.J., A comparison of three instructional approaches for teaching written composition to high school juniors: teacher lecture, peer evaluation, and group tutoring. Diss. Boston University School. Boston, 1977.

Fellows, J.E., The influence of Theme-Reading and Theme-Correction on Eliminating Technical errors in the Written Compositions of Ninth Grade Pupils. In: Studies in Education, deel VII. University of Iowa. Z.p., 1932.

Feteris, E.S.F. en N. Lewis, Over en weer (4 V.W.O.). Meulenhoff
Educatief. Amsterdam, 1982.

Flanigan, M.C. and D.S. Menendez, Perception and Change: Teaching revision. In: College English, jrg.42, no.3, 1980, p.256-266.

Flavell, J.H., Cognitive development. Prentice-Hall. Englewood Cliffs, N.J., 1977.

Flower, L. & J.R. Hayes, A cognitive process theory of writing. In: College Composition and Communication, jrg.32, no.4 (dec.), 1981, p.365-387.

Flower, L., Writer-based prose: a cognitive basis for problems in writing. In: College English, jrg.41, no.1 (sept.), 1979, p.19-37.

Flynn, E.A., Freedom, restraint, and peer group interaction. Paper Annual Meeting of the Conference on College Composition and Communication. San Fransisco, 1982. (ED 216 365).

Ford, B.W., The effects of peer editing/grading on the grammar-usage and theme-composition ability of college freshman. Diss. University of Oklahoma. Z.p., 1973.

Forman, J., Notes toward writing across the curriculum: Some collaborative efforts. In: Journal of Basic Writing, jrg.2, no.4, 1980, p.12-21.

Fox, R.F., Treatment of writing apprehension and its effects on composition. Diss. University of Missouri-Columbia. Z.p., 1978. (UMI 79-15233).

Fox, R.F., Treatment of writing apprehension and its effects on composition. In: Research in the teaching of English, jrg.14, no.1, (feb.), 1980, p.39-49.

Freed, P., Writing with the opaque projector. In: English Journal, jrg.70, no.2, 1981, p.31-33.

Gebhardt, R., Teamwork and feedback: Broadening the base of collaborative writing. In: College English, jrg.42, no.1, 1980, p.69-74. (ED 174 994).

Glopper, K. de, Schrijfprestaties in het voortgezet onderwijs. Een onderzoek naar de stelvaardigheid van leerlingen in het derde leerjaar L.T.O., L.H.N.O., M.A.V.O., H.A.V.O. en V.W.O. S.C.O., S.V.O. (Selectareeks). Amsterdam, 1985a.

Glopper, K. de, Verhalen, betogen en beschouwen. Voorbeelden van schrijfprestaties van leerlingen in het derde leerjaar van het voortgezet onderwijs. S.C.O.-rapport 60. S.C.O. Amsterdam, 1985b.

Glopper, K. de, Functionele schrijfopdrachten. Voorbeelden van schrijfprestaties van leerlingen in het derde leerjaar van het voorgezet onderwijs. S.C.O.-rapport 61. S.C.O. Amsterdam, 1985c.

Glucksberg, S., R. Kraus and E.T. Higgins, The development of referential communication skills. In: F. Horowitz (ed.), Review of child development research, deel 4. University of Chicago Press. Chicago, 1975, p. 305-345.

Goldsmith, E., Peer editing with guidelines. In: Hayes, M.F., M. Morenberg en J. Ziegler, Teachers at work: articles from the Ohio Writing Project. Summer Institute Miami University. Oxford (Ohio), 1982, p. 46-53. (ED 232 209).

Golsby, J., Peer tutoring in basic writing: a tutor's journal. Classroom research study no.4. Berkeley, 1981. (ED 250 717).

Gregg, L.W. and E.R. Steinberg (eds.), Cognitive processes in writing. Lawrence Erlbaum Associates. Hillsdale, N.J., 1980.

Griffioen, J. en H. Damsma, Zeggenschap. Grondslagen en een uitwerking van een didactiek van het Nederlands in het voortgezet onderwijs. Tweede, herziene en vermeerderde druk. Wolters-Noordhoff. Groningen. Z.j. (1978).

Griffioen, J., m.m.v. H. Damsma en A.J. Evenboer, Tegenspraak. Handelingsoriëntaties voor de leraar Nederlands in het voortgezet (basis)onderwijs. Wolters-Noordhoff. Groningen, 1982.

Grift, W. van de, Het innovatieprobleem en het gedrag van schoolleiders. Diss. R.U.L. (te verschijnen).

Groot, A.D. de, Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschap. Mouton.'s-Gravenhage, z.j. (1972).

Gross, G.J., Work together. Write together: Group projects in technical writing. Paper presented at the Annual Meeting of the Conference on College Composition and Communication (28th, March 31-April 2). Kansas City, 1977. (ED 146 602).

Gwyn, C. and D. Swanson-Owens, Essay editing: Helping students teach themselves. Z.p., 1980. (ED 192327).

Hafernik, J.J., The how and why of peer editing in the ESL writing class. In: CATESOL Occasional papers, nr.10, z.p., 1984, p.48-58. (ED 253 064).

Hagen, A., Standaard en dialect-sprekende kinderen, een studie over monitoring in het taalgebruik. Coutinho. Muiderberg, 1981.

Hansen, S.S. en L. Vogt, A variation in peer critiquing: peer editing as the integration of language skills. In: Kwalick, B. (ed.), Selected papers from the 1982 Conference "New York Writes: Kindergarten through College". City University of New York. New York, 1982, p.34-37. (ED 234 401).

Hawkins, T., Training peer tutors in the art of teaching. In: College English, jrg.40, no.4, 1978, p.440-443.

Hayes, J.R. and L.S. Flower, A cognitive model of the writing proces in adults. Final report. Carnegie Mellon University. Pittsburg, 1983. (ED 240 608).

Hayes, J.R. and L.S. Flower, Identifying the organization of writing processes. In: L.W.Gregg en E.R.Steinberg (eds.) Cognitive processes in writing. Lawrence Erlbaum Associates. Hillsdale, N.J.,1980, p.3-30.

Healy, M.K., Using student writing response groups in the classroom. Bay Area Writing Project Curriculum Publication no.12. University of California. Berkeley, 1980.

Hilgers, T., Toward a taxonomy of beginning writers' evaluative statements on written compositions. In: Written Communication, jr.1, no.3, (jul.), 1984, p.365-384.

Hoogeveen, M. en M. Verkampen, Schrijfonderwijs in praktijk. Een verslag van een etnografisch onderzoek naar de invoering van thematisch-cursorisch schrijfonderwijs op een basisschool. S.L.O. Enschede, 1985.

Hoover, R.M., Experiments in peer teaching. In: College Composition and Communication, jrg.23, 1972, p.421-428.

Houben, H., Correctie-euvel. In: Levende talen, 1930, p.25-29.

House, W.C., Effects of knowledge that attributions will be observed by others. In: Journal of research in personality, jrg.14, no.14, 1980, p.528-545.

Howgate, L., Building Self-Esteem through the writing process. Writing teachers at work. Berkeley, 1982. (ED 250 716).

Humes, A., Research on the Composing process. In: Review of Educational Research, jrg.53, no.2 (summer), 1983, p.201-216.

Hurlow, M., Alleviating writing anxiety in individual and small-group settings. Paper Annual Meeting of the Writing Centers Association. West-Lafayette, 1983. (ED 238 008).

Irwin, D.E., Effects of peer tutoring on academic achievement and affective adjustment. In: Enright, G., (ed.). The 1980's: New Sources of energy for learning. Proceedings of the Annual Conference of the Western College Reading Association (13th). San Fransisco, 1980, p.42-46. (ED 247 536).

Jager, B., J.L. van der Velde en B. Slettenaar, Taalgoed. (3vh). Thieme. Zutphen, 1983.

James, D.R., Peer teaching in the writing classroom. In: English Journal, jrg.70, no.7, 1981, p.48-50.

Jansen, F. en A. Scheringa, Doel- en publiekgerichte opstellen. Verslag van een literatuurstudie naar kenmerken van doel- en publiekgerichte persuasieve teksten. Projectpublikatie 5 van Z.W.O.-project Onderzoek naar effecten van beoordeling door medeleerlingen op aspecten van de stelvaardigheid. S.C.O. Amsterdam, 1984.

Jones, T.L, An exploratory study of the impact on student writing of peer evaluation. Diss. Michigan State University. Z.p., 1977. (UMI 78-10076).

Jones, W., Revision for basic writers. Paper presented at the Annual Meeting of the Conference on College Composition and Communication (32nd, March 26-28). Dallas, 1981. (ED 202 020).

Karegianes, M.L., E.T.Pascarella and S.W.Pflaum, The effects of peer editing on the writing proficiency of low-achieving tenth grade students. In: Journal of Educational Research, 73, 1980, p.203-207.

Kish, L., Survey sampling. Wiley ans Sons. New York, 1965.

Korrektie-(K)euvel In: Berichten en mededeelingen van de vereeniging van leeraren in levende talen. 1929, p. 24-28.

Kroon, J.W.M., Grammatica en communicatie in het onderwijs Nederlands. Diss. K.U.N. Groningen, z.j. (1985).

Lagana, J.R., The development, implementation, and evaluation of a model for teaching composition which utilizes individualized learning and peer grouping. Diss. University of Pittsburg. Pittsburg, 1972.

Lamberg, W., Self-provided and peer-provided feedback. In: College Composition and Communication, jrg.31, no.1, 1980, p.63-69.

Laney, J.D. Composition in the intermediate grades: How to promote thinking and creativity. Paper Annual meeting California Educational Research Association. Los Angeles, 1983. (ED 241 938).

Langer, J.A. and A.N. Applebee, Learning to manage the writing process: tasks and strategies. Z.p., 1983. (ED 234 420).

Laque, C.F. and P.A.Sherwood, A Laboratory Approach to Writing. Urbana, 1977.

Leidse Werkgroep Moedertaaldidactiek, Moedertaaldidactiek. Een handleiding voor het voortgezet onderwijs. Coutinho. Muiderberg, 1980.

Lewis, U.E., Peer-evaluation in a writing seminar. Ohio Wesleyan University. Delaware, 1981. (ED 226 355).

Lutkus, A.D., Using PEERRATE: A computerized system of student term paper grading. Paper presented at the Annual Convention of the American Psychological Association Toronto 86th, 1978. (ED 167 887).

Lyons, W.D., The effects of teacher-peer response and teacher-only response upon attitudes toward writing, and upon writing performance. Diss. University of Missouri-Colombia. Z.p., 1976.

Maimon, E.P., Talking to strangers. In: College Composition and Communication, jrg.30, no.4, 1979, p.364-369.

Maize, R.C., Two methods of teaching English composition to retarded college freshmen. In: Journal of Educational Psychology, 45, (Jan.) 1954, p.22-28.

Manzo, A.V. and J.K.Sherk, Manzo's Language Shaping Paradigm (LSP). Z.p., 1977. (ED 181 460).

Martin, W.R., Evaluating on the right side. Paper Annual Meeting of the NCTE. Denver, 1983. (ED 241 939).

Mazurek, K., Idea exchange: A method of marking to reduce drudgery and increase learning. In: English Quarterly, jrg.12, no.4, 1979, p.102-105.

Megna, J.F., Teaching writing skills through peer evaluation. In: The English Record, 1976, p.98-106.

Mellenbergh, G.J., Bekend, maar onbemind (deel 1). Subfaculteit Psychologie Universiteit van Amsterdam. Amsterdam, 1976.

Mellenbergh, G.J., Bekend, maar onbemind (deel 2): Experimentele proefopzetten. Subfaculteit Psychologie Universiteit van Amsterdam. Amsterdam, 1977.

Meuffels, H.L.M., Studies over taalvaardigheid. Diss. U.v.A. Amsterdam, 1982.

Meijerink, H. en J. Vos, Tussen de regels. (Deel 2). Wolters-Noordhoff. Groningen, 1978.

Miller, M.D. and J.A. Daly, The development of a measure of writing apprehension. Chicago, 1975. (Educational Testing Service: Tests in Microfiche, no.006 172).

Moffett, J., Teaching the universe of discourse. Houghton Mifflin Company. Boston, 1968.

Myers, C.F., Teacher and peer evaluative feedback in the development of two composition skills: punctuation and paragraph unity. Diss. St.John's University. New York, 1979.

Newkirk, T., How students read student papers: an exploratory study. In: Written Communication, jrg.1, no.3, (jul.), 1984a, p.283-305.

Newkirk, T., Direction and misdirection in peer response. In: College Composition and Communication, jrg.35, no.3, (oct.), 1984b, p.301-311.

Nijmeegse Werkgroep Taaldidactiek, Taaldidactiek aan de basis. Wolters-Noordhoff. Groningen, z.j. (1978).

O'Donnell, C., Peer Editing: A way to improve writing. Paper presented at the combined Annual Meeting of the Secondary School English Conference and the Conference on English Education (March, 27-29). Omaha, 1980. (ED 189 604).

Osborn, L., Language arts: Writing families. In: Instructor, jrg.90, no.2, 1980, p.152-155.

Parks, K.D., A key to clarity - some variations on a slide-tape extravaganza. Paper presented at the Annual Meeting of the Conference on College Composition and Communication (28th, March 31-April 2). Kansas City, 1977. (ED 145 448).

Pasternack, S.R., Properly motivated. Students become good peer graders. In: Journalism Educator, jrg.36, no.3, 1981, p.18-18.

Peckham, I., Peer Evaluation: Close-up: Evaluating and revising composition. In: English Journal, jrg.67, no.7, 1978, p.61-63.

Pfeifer, J.K., The effects of peer evaluation and personality on writing anxiety and writing performance in college freshmen. Diss. Texas Tech. University. Z.p., 1981.

Pianko, S. and A. Radzik, The student editing method. In: Theory into Practice, jrg.19, no.3, 1980, p.220-224.

Pierson, H., Peer and teacher correction: A comparison of the effects of two methods of teaching composition in grade nine English classes. Diss. New York University. New York, 1966.

Plevin, A., The basic comp. class as rhetorical situation. Paper Annual meeting of the Conference on College Composition and Communication. San Fransisco, 1982. (ED 216 347).

Popham, R.J. and J.Zarem, Improving basic writing skills. Instructional objectives exchange. Los Angeles, 1978.

Reichardt, C.S., The statistical analysis of data from nonequivalent group designs. In: T.C. Cook en D.T. Campbell, Quasi-Experimentation. Designs & analysis issues for field settings. Houghton Miffin Company. Boston, z.j. (1979), p.147-205.

Reid, L., Talking: the neglected part of the writing process. Paper Annual Meeting of the National Council of Teachers of English. Seattle, 1983. (ED 229 762).

Reigstad, T.J. and D.A. McAndrew, Training tutors for writing conferences. NCTE. Urbana, 1984. (ED 240 589).

Rivera-Hernandez, N., Collaborative Learning: a means of improving students' exam writing. Paper National Endowment for the Humanities. Glenside, 1982. (ED 221 861).

Roundy, N., Revision Pedagogy in technical writing. Paper Annual Meeting of the Conference on College Composition and Communication. New York, 1984. (ED 243 119).

Rubin, D., Once is not enough. Paper Annual meeting Conference on College Composition and Communication (March). Detroit, 1983.

Rubin, D., Evaluating freshman writers: what do students really learn? In: College English, jrg.45, no.4 (april), 1983, p. 373-379.

Rijlaarsdam, G., Leerlingenrespons: een kansrijke didactiek van het stellen. Een verslag van een literatuuronderzoek naar de effecten van leerlingenrespons (peer evaluation) op stelvaardigheid. S.C.O.-rapport 27. S.C.O. Amsterdam, 1984.

Rijlaarsdam, G. en H. Blok, Beoordeling van schrijfprodukten door leerlingen: theorie en praktijk. In: Levende talen, nr. 365 (oktober), 1981, p.753-766.

Rijlaarsdam, G., Wat leren leerlingen volgens eigen zeggen? Effecten van leerlingenrespons, gemeten aan zelfrapportages. Lezing Congres Vereniging Interuniversitair Overleg Taalbeheersing. Tilburg, aug. 1984. Verschenen in: H. Blok e.a., Moedertaalonderwijs in onderzoek. S.C.O.-cahier 28. S.C.O. Amsterdam, 1984, p.17-32. Eveneens verschenen in Moer, 1, 1985, p.24-32 en in W.K.B. Koning (red.), Taalbeheersing in theorie en praktijk. Lezingen van het V.I.O.T.-taalbeheersingscongres, gehouden op 28, 29 en 30 augustus 1984 aan de Katholieke Hogeschool te Tilburg. Foris Publications. Dordrecht, 1985, p.429-438.

Rijlaarsdam, G. and H. Hulshof, Schrijfonderwijs: overwegingen voor een leerplan. In: G. Rijlaarsdam en H. Hulshof (red.), Schrijven: theorie en praktijk van het schrijfvaardigheidsonderwijs. D.C.N.-cahier 16. Malmberg. Den Bosch, z.j. (1984), p.173-192.

Rijlaarsdam, G., W. Groeneweg en T. Weyers, Een didactiek voor het ontwikkelen van communicatief inzicht: reageren op elkaars opstellen. In: Levende talen, nr.387, (themanummer Communicatief Taalonderwijs), (dec.) 1983, p.622-625.

Sager, C., Improving the quality of written composition through pupil use of rating scale. Paper presented at the Annual Meeting of the National Council of Teachers in English. Z.p., 1973. (ED 089 304).

Sager, C., Improving the quality of written composition through pupil use of rating scale. Diss. Boston University. Boston, 1972.

Salili, F., M.L. Maehr, R.L. Sorensen and L.J. Fyans jr., A further consideration of the effects of evaluation on motivation. In: American Educational Research Journal, jrg.13, no.2, 1976, p.85-102.

Sarbin, T.R. and V.L. Allen, Role theory. In: G. Lindzey and E. Aronson (eds.), Handbook of social psychology. Deel 1. Reading, 1968, p.488-567.

Sarbin, T.R., Cross-age tutoring and social identity. In: V.L. Allen (ed.), Children as teachers. Theory and research. Academic Press. New York, 1976, p.27-40.

Scardamalia, M. and C. Bereiter, Research on written composition. In: M.C. Wittrock (ed.), Third Handbook of research on teaching. Macmillan. New York, 1986, p.778-803.

Schneiders, A.J., Correctie-euvel. In: Levende talen, 1930, p.16-18.

Schrader, M.A. and R. Westphal, Reliability of student evaluations of students performance. Final Report. Cleveland, 1976. (ED 134 622).

Schrader, M.A. and R.Westphal, Reliability of student evaluations of student performance. Final Report. Cleveland, 1976. (ED 134 622).

Schulz van Thun, F., Hoe bedoelt u? Wolters-Noordhoff. Groningen, 1982.

Schuster, C.I., The unassignment: writing groups for advanced expository writers. Paper Annual Meeting of the Wyoming Conference of Freshman and Sophomore English. Laramie, 1983. (ED 234 409).

Sears, M.O., Effects of a student centered procedure on the self-concepts and writing practice of college freshmen. Diss. Florida State University. Z.p., 1970.

Sears, P., Letter writing: A technique for treating writer's block. Paper presented at the Annual Meeting of the Conference on College Compositon and Communication (30th, April 5-7). Minneapolis, 1979. (ED 176 337).

Selfe, C.L., Using groups to pre-evaluate papers in the technical writing classroom. Michigan, 1981. (ED 226 369).

Shantz, C.U., The development of social cognition. In: Review of child development research. Deel 5. University of Chicago Press. Chicago, 1975, p.257-323.

Sherwood, J.J., Increased self-evaluation as a function of ambiguous evaluations by referent others. In: Sociometry, jrg.30, no.4, 1967, p.404-409.

Silver, M., Using peer critiques to train peer tutors. In: College English, jrg.40, no.4, 1978, p.433-436.

Silver, S.E. and B.R. Schlenker, Self-presentational reactions to interpersonal evaluations. Paper presented at the Annual Convention of the American Psychological Association (89th, Los Angeles, August 24-26). Los Angeles, 1981. (ED 210 586).

Smelstor, M., A guide to evaluating students' writing. Wisconsin University. Madison, 1978. (ED 176 275).

Smith. M.E., Peer tutoring in a writing workshop. Diss. University of Michigan. Z.p., 1975. (U.M.I. 76 9320).

Smith, E., Evaluation of writing with semantic differential scales. Paper Annual Meeting of the Northwest Regional Conference of the NCTE. Boise, 1982. (ED 222 900).

Smith, E., Student writing in the foreground of a literature class. Paper Annual Meeting of the National Council of Teachers of English. Denver, 1983. (ED 239 261).

Snipes, W.C., An inquiry. Peer group teaching in freshmen writing. In: College composition & communication, 22, 1971, p.1969-ev.

Soares, A.T. and L.M. Soares, The gender differential in interpersonal perception. Paper presented at the Annual Meeting of the Eastern Psychological Association (51th, April 9-12). Hartford, 1980. (ED 190 945).

Soven, M., The research paper: the student as audience. Z.p., 1980. (ED 218 652).

Spigelmire, L., A peer review process for developmental writers. In: Exercise Exchange, jrg.26, no.1, 1981, p.20-23.

Spina, L. and D. Welhoelter, Writing bugs become reading bugs. Paper presented at the Annual Meeting of the Missouri State Council of the International Reading Association (14th, March 13-14). Columbia, 1981. (ED 205 990).

Steinacher, R.C., Learning to write by writing in a community college remedial composition course. Diss. Florida State University. Z.p., 1976. (U.M.I. 76 28, 641).

Stone, W.B., Rewriting in advanced composition. Paper Annual meeting of the Conference on College Composition and Communication (32nd, March 26-28). Dallas, 1981.

Swarts, H., L.S. Flower and J.R. Hayes, Designing Protocol Studies of the writing process: an introduction. Carnegie Mellon University. Pittsburg, 1980.

Tremmel, R., The world and the writing classroom: the teaching of the writers-in-the-schools. Paper Annual Meeting of the Conference on College Composition and Communication. Detroit, 1983.

Triesscheyn, B., I. Bochardt en G. Rijlaarsdam, Wat vind je van mijn opstel? Leerlingen becommentariëren elkaars opstellen: praktijkervaringen van leerlingen en leraren. Projectpublikatie nr.1. Z.W.O.-project 'Onderzoek naar effecten van beoordeling door mede-leerlingen op aspecten van de stelvaardigheid'. S.C.O. Amsterdam, 1984.

Turbill, J., Now, we want to write! Primary English Teaching Association. Rozelle, 1983. (ED 240 590).

Tuyl, H. van, Etnografisch onderzoek, retoriek en communicatief taalonderwijs. In: L. Lentz, J. Sturm en H. van Tuyl (red.): Schrijfonderwijs onder het mes. S.L.O. Enschede, 1986.

Verslag van de adviesvoorbereidende studiedag schrijfonderwijs H.A.V.O./V.W.O. (17 juni 1984). Adviescommissie voor de Leerplanontwikkeling Moedertaal i.o.. Z.p., z.j. (Enschede, 1985).

Verslag van de adviesvoorvereidende studiedag voor docenten M.A.V.O./L.B.O. (6 juni 1985). Adviescommissie voor de Leerplanontwikkeling Moedertaal i.o. Z.p. (Enschede), 1985.

Verstegen, G., Twaalf leraarsdagboeken over schrijven. Verslag en literatuurverkenning. Adviescommissie voor de Leerplanontwikkeling Moedertaal i.o. Z.p. (Enschede), 1985.

Vlasman, C. (m.m.v. A.G. van Balen), Aantekeningen bij De papieren wereld, Deel 1 en Deel 2. (Katholiek Pedagogisch Centrum). Z.p., z.j.('s-Hertogenbosch, 1979).

Vos, J. en H. van der Laan, Taal vandaag. (Deel 1). Docentenhandleiding. Ten Brink. Meppel, 1978.

Vos, J. en H. van der Laan, Taal vandaag. (3). Ten Brink. Meppel, z.j.

Vijn, P., De analyse van kruistabellen. In: Tijdschrift voor onderwijsresearch, jrg. 7, no.5, 1982, p.212-220.

Wagner, E.N., How to avoid grading compositions. In: English Journal, jrg.1964, no.3, 1975, p.76-79.

Wagner, E.N., The Impact of Composition Grading on the Attitudes and Writing of Freshman English Students. Diss. University of Virginia. Z.p., 1975.

Ward, L.R., A comparison of two methods of teaching writing in a course in communication skills. Diss. Purdue University. Z.p.,1959.

Warner, M.P., A Successful writing program for High School learning. In: Pointer, jrg.23, no.2, 1979, p.30-36.

Weeks, J.O. and E.M. White, Peer editing versus teacher editing: does it make a difference? Paper Meeting of the North Carolina Council of the International Reading Association. Charlotte, 1982. (ED 224 014).

Weiner, B., et al., Perceiving the causes of success and failure. In: E.E. Jones e.a. (eds.), Attribution: perceiving the causes of behavior. General Learning Press. Morristown, (N.J.), 1972.

Wesdorp, H., Het meten van produktief-schriftelijke vaardigheid. Directe en indirecte methoden 'opstelbeoordeling' versus 'schrijfvaardigheidstoetsen' (enigszins gewijzigde versie diss. U.v.A.). Muusses. Purmerend, 1974.

Wesdorp, H., Evaluatietechnieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid. (S.V.O.-reeks no.42). Staatsuitgeverij. 's-Gravenhage, 1981.

Wesdorp, H., Schrijven in het voortgezet onderwijs. Een overzicht van het onderzoek naar de effecten van diverse instructie-variabelen op de stelvaardigheid. (S.V.O.-reeks no.71). S.V.O./Flevodruk Harlingen. Z.p. ('s-Gravenhage), 1983.

Wesdorp, H., Goed onderwijs, wat is dat? Voorstudie periodieke peiling van het onderwijsniveau. Een samenvatting van de resultaten van het proefonderzoek naar het taalonderwijs in de zesde klas van de basisscholen. (Selectareeks). S.V.O., S.C.O., O. & W. Z.p. ('s-Gravenhage), z.j. (1985).

Wesdorp, H. en J.B. Hoeksma, Voorstudie Periodieke Peiling van het Onderwijsniveau. Deel I: Leer- en schrijfprestaties van zesde klassers. S.C.O.-rapport 55, S.C.O. Amsterdam, 1985.

White, K., The effect of source of evaluation on the development of internal control among young boys. In: Psychology in the Schools, jrg.9, no.1, 1972, p.56-61.

Wouters, L. en T. de Jong, Hardop denken tijdens tekstbestudering. In: Tijdschrift voor Onderwijsresearch, jrg. 7, no.2, (maart), 1982, p.60-75.

Yeager, J.L., C. Wood and M.A. Kissel, An analysis of pupil activities in an IPI classroom. In: C.M. Lindvall and R.C. Cox, Evaluation on a toll in curriculum development: the I.P.I. evaluation program. Rand McNally. Chicago, 1970.

Zanger, J. de, Praktijkboek Gericht Schrijven. V.W.O./ H.A.V.O. Deel 1. S.L.O. Enschede, z.j. (1981).

Zanger, J. de, Praktijkboek Gericht Schrijven. V.W.O./ H.A.V.O. Deel 2. S.L.O. Enschede, z.j. (1981).

Ziv, N.D., Peer groups in the composition classroom: a case study. Paper Annual meeting of the Conference on College Composition and Communication (34th, March 17-19). Detroit, 1983. (ED 229 799).

Bijlage 1: Geraadpleegde bronnen, gebruikt bij het ontwerp van een theorie over leerlingenrespons (par. 2.2.)

Barry, 1980
Bean, 1979
Beaven, 1977
Beck, 1978
Bell, 1983
Benesch, 1984
Berkenkotter, 1983
Beyer and Brostoff, 1979
Bissland, 1980
Blake and Tuttle, 1977
Booher, 1982
Bruffee, 1980
Buys, 1984
Calkins, 1978
Calzonetti, 1981
Camplese and Mayo, 1982
Christensen, 1977
Clifton, 1980
Coleman, 1978
Collins, 1983
Covington, 1979
Crews, 1983
Crowhurst, 1979
Damsma, 1985
Danis, 1982
Elbow, 1973
Elias, 1982
Ellman, 1975
Engel, 1983
Flanigan and Menendez, 1980
Flynn, 1982
Forman, 1980
Freed, 1981
Gebhardt, 1980
Goldsmith, 1982
Golsby, 1981
Griffioen en Damsma, 1978
Griffioen e.a., 1982
Gross, 1977
Gwyn and Swanson-Owens, 1980
Hafernik, 1984
Hansen and Vogt, 1982
Hawkins, 1978
Healy, 1980
Hoover, 1972
Howgate, 1982
Hurlow, 1983
Irwin, 1980
James, 1981
Jones, 1981
Lamberg, 1980
Laney, 1983
Langer and Applebee, 1983
Laque, 1977
Leidse Werkgroep Moeder-moedertaaldidactiek, 1980
Lewis, 1981
Lutkus, 1978
Maimon, 1979
Manzo and Sherk, 1977
Martin, 1983
Mazurek, 1979
Megna, 1976
Moffett, 1968
Nijmeegse Werkgroep Taal-didactiek, 1978
O'Donnell, 1980
Osborn, 1980
Parks, 1977
Pasternack, 1981
Peckham, 1978
Pianko and Radzik, 1980
Plevin, 1982
Popham and Zarem, 1978
Reid, 1983
Reigstad and McAndrew, 1984
Rivera-Hernandez, 1982
Roundy, 1984
Rijlaarsdam en Blok, 1981
Sager, 1973
Schuster, 1983
Sears, 1979
Selfe, 1981
Silver 1978
Smelstor, 1978
Smith, 1975
Smith, 1982
Smith, 1983
Snipes, 1971
Soven, 1980
Spigelmire, 1981
Spina and Welhoelter, 1981
Steinacher, 1976
Straver, 1981
Tremmel, 1983
Turbill, 1983
Wagner, 1975
Warner, 1979
Weeks and White, 1982

Bijlage 2: Geraadpleegde effectstudies (par. 3.3.)

1. Benson (1979)
2. Bouton and Tutty (1975)
3. Burt (1980)
4. Carter (1982)
5. Clifford (1981)
6. Copland (1980)
7. Delaney (1980)
8. Earls (1983)
9. Farrell (1977)
10. Ford (1973)
11. Fox (1978)
12. Karegianes et al. (1980)
13. Lagana (1972)
14. Lyons (1976)
15. Maize (1954)
16. Myers (1979)
17. Pfeifer (1981)
18. Pierson (1967)
19. Sager (1973)
20. Sears (1971)
21. Ward (1959)

## Bijlage 3: Lespraktijken op tien scholen in Nederland met leerlingenrespons (par.4.1.)

| | 1. LHNO | 2. MAVO | 3. MAVO | 4. MAVO | 5. HAVO/VWO |
|---|---|---|---|---|---|
| **OPDRACHTEN** | leraar kiest 6 à 7 zakelijke en verhalende opdrachten; publiek: niet gedefinieerd; op school beginnen en thuis afmaken | leraar kiest 10 à 15 examentitels, verhalend en betogend; publiek: niet gedefinieerd; op school (2 lesuren) thuis afmaken | leraar geeft titels op: actuele, verhalende en beschouwende; publiek: niet gedefinieerd; op school beginnen en thuis afmaken | leraar geeft opdracht; publiek: de leraar op school (2 lesuren) thuis afmaken | mini cyclus bij B.2 opstel lln. kunnen kiezen uit 3 door leraar voorbereide onderwerpen; publiek: niet gedefinieerd op school voorbereiding en thuis schrijven |
| **AANTAL** | 7 à 8 per jaar | 6 keer per jaar | 8 keer per jaar | 3 à 4 per jaar | In havo/vwo-4: 2 à 3 per jaar havo/vwo-5: 1 per jaar vwo-6: 1 à 2 per jaar |
| **VOORBEREIDING** | 20 min. instructie voorafgaande aan de opdracht | geen | geen | 2 à 3 lesuren met voorbereidende oefeningen | er wordt in groepjes gewerkt lln. maken uittreksels van artikelen en wisselen deze uit. Vervolgens wordt gezamenlijk een opstel-schema gemaakt. |
| **LEERLINGEN-REPONS** | niet altijd | altijd, 3 leerlingen lezen hun opstel voor | altijd, iedere leerling krijgt het schrift van een ander | altijd: "stapel commentaar": 1 ll. geeft 1e commentaar, 2e leerling vermeldt of deze het ermee eens is, 3e leerling vermeldt of hij het met 1e of 2e leerl. eens is. | altijd: elke leerling leest een opstel over een ander onderwerp dan waarover hij zelf geschreven heeft. Leerlingen besteden één lesuur aan het becommentariëren |
| **COMMENTAAR GEVEN** richtlijnen | geen | ja ("Opbouw"-methode) | ja, gedicteerd in schrift (eigen fabrikaat) | ja, oefenopdrachtjes (eigen fabrikaat) | ja, lijstje met punten (eigen fabrikaat) |
| Voorlezen | soms, 2 à 3 keer per jaar | ja | nee | ja, de oefenopdrachtjes | nee |
| Hoeveel commentatoren | 1 per opstel, bij voorlezen alle leerlingen | alle andere leerlingen kunnen mondeling reageren | 1 per opstel | 3 à 4 per opstel | 1 per opstel |
| Wijze van commentaar geven | mondeling | mondeling | schriftelijk, samenwerkend | schriftelijk | schriftelijk, moet minstens 1 foliokantje beslaan |
| **WAT DOET LERAAR MET LEERLINGEN-COMMENTAAR** | beoordeelt onafhankelijk van leerlingen commentaar | beoordeelt 3 opstellen per keer,oordeel van lln. wordt veelal niet in eindoordeel betrokken | beoordeelt na de comm.-ronde 5 of 6 opstellen per keer,controleert of overal comm. bij staat, rekent verbeteringen v. mede-lln. niet bij de fouten mee. Beoordeelt het commentaar | beoordeelt alle opstellen. Vergelijkt zijn commentaar met leerlingencommentaar | leraar beoordeelt per groepje eerst de opstellen en dan het leerlingen commentaar. Daarna wordt de schrijfmap in de klas besproken. |
| **VERWERKEN VAN COMMENTAAR** | geen herschrijvingen | leerlingen verbeteren eerst fouten en lezen dan voor | geen herschrijvingen | leerlingen moeten het opstel nog in het net schrijven; verwerken van comm. (verbeteren, herschrijven) mág, maar moet niet | geen herschrijvingen. Voordat leerlingen aan een nieuw opstel beginnen worden ze geacht eerst hun oude opstellen door te lezen |

(vervolg bijlage 3)

| | 6. HAVO/VWO | 7. HAVO/VWO | 8. HAVO/VWO | 9. HAVO/VWO | 10. HAVO/VWO |
|---|---|---|---|---|---|
| OPDRACHTEN | schrijfopdrachten uit de methode 'Taalwerk' of door de leraar gemaakt; publiek: niet gedefinieerd; op school beginnen en thuis afmaken | onderwerp wordt van te voren door leraar gegeven; zakelijk of verhalend; zelf gekozen publiek deelopdrachten: afwisselend op school of thuis geschreven | leerlingen mogen zelf een titel kiezen, kladversie en schema op school maken; publiek: niet gedefinieerd; oefenopdrachten en opstel-afwisselend thuis of op school maken | leraar maakt de opdrachten zelf. publiek: klasgenoten of naar keuze; zowel thuis als op school schrijven | leerlingen kunnen kiezen uit 10 onderwerpen. Vaak examenopdrachten van voorgaande jaren. Zowel betogend als verhalend; Publiek: niet gedefinieerd; opstel op school schrijven, oefeningen soms ook thuis |
| AANTAL | 3 à 4 per jaar | 2 à 3 per jaar | gemiddeld 7 opstellen vanaf de 3e klas: 1 informerend, 3 verhalende en 3 betogende | 3 opstellen per jaar | 3 opstellen per jaar |
| VOORBEREIDING | verschillende schrijfopdrachten gaan vooraf aan het uiteindelijke opstel, materiaal mag gebruikt worden (informatie, woordenboek) | per opstel 2 à 3 lessen voorbereidende opdrachten | uitgebreide voorbereidende oefeningen a.d.h.v. 'De Papieren Wereld' aantal voorbereidende lessen afhankelijk van type opstel | per opstel een aantal voorbereidende lessen | per opstel een aantal voorbereidende lessen |
| LEERLINGEN-RESPONS | altijd, op verschillende manieren | altijd bij toetsopstellen. Meestal bij deelopdrachten | lezen: altijd. Becommentariëren is afhankelijk van de tijd, klassikaal of in groepjes | altijd: er wordt in groepjes van 3 à 4 leerlingen gewerkt | soms, op ongestructureerde wijze |
| COMMENTAAR GEVEN | | | | | |
| Richtlijnen | ja, lijstje met punten uit 'Taalwerk' | ja, beoordelingsschema uit 'Over en Weer' | ja, beoordelingsschema uit 'De Papieren Wereld' | ja, door leraar gegeven | geen |
| Voorlezen | soms | soms afhankelijk van de opdracht: zelfselectie | soms, elk groepje kiest het beste opstel uit | ja, één per groepje | soms, niet verplicht |
| Hoeveel commentatoren | soms één, soms meer | één of twee | minimaal drie | drie per opstel | - |
| Wijze van commentaar geven | afwisselend mondeling of schriftelijk. Soms in de vorm van cijfers, soms in de vorm van een waarde-oordeel | schriftelijk, leerlingen zetten het commentaar met potlood onder de tekst en verbeteren fouten | mondeling of schriftelijk, leerlingen geven punten per beoordelingsaspect | eerst binnen een groepje mondeling en schriftelijk, daarna een keuze uit de opstellen klassikaal. Af en toe wordt gewerkt met schaalpunten (1 t/m 5) | leerlingen lezen elkaars opstellen soms en geven elkaar mondeling commentaar |
| WAT DOET LERAAR MET LEERLINGEN COMMENTAAR | leraar beoordeelt de uiteindelijke opstellen, geeft commentaar en een cijfer | | nadat de leerlingen elkaar becommentarieerd hebben, kijkt de leraar de opstellen na. Hij geeft commentaar en een cijfer. Leerlingen comm. legt hij terzijde | leraar beoordeelt opstellen, in de derde klas geeft hij nog geen cijfer | leraar beoordeelt de opstellen, hij onderstreept fouten en plaatst opmerkingen in de kantlijn of onder het opstel. Daarna klassikale bespreking |
| VERWERKEN VAN COMMENTAAR | geen herschrijvingen punten van kritiek verbeteren in volgende schrijfopdracht | geen herschrijvingen op grond van commentaar | aan de hand van het commentaar mogen leerlingen correcties aanbrengen, het opstel verbeteren of evt. opnieuw [illegible]rijven | geen herschrijving | geen herschrijving |